

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:31:50 ; Search time 246 Seconds  
(without alignments)

8651.818 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325

Perfect score: 501

Sequence: 1 atgtttacacgtcgggcac.....cccgctcgaggctcgctgc 501

Scoring table: OLIGO NUC

Gapop\_60.0 , Gapext 60.0

Searched: 3373863 seqs, 212409041 residues

Word size : 15

Total number of hits satisfying chosen parameters: 1847

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001as:\*
- 5: Geneseq2001bs:\*
- 6: Geneseq2002s:\*
- 7: Geneseq2003as:\*
- 8: Geneseq2003bs:\*
- 9: Geneseq2003cs:\*
- 10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	501	100.0	53789	2 AAV21187	AAV21187 Amycolato
2	26	5.2	14055	7 AAL61170	AAL61170 Actinosyn
3	26	5.2	82746	7 AAL61224	AAL61224 Actinosyn
4	26	5.2	88421	6 AAL40781	AAL40781 842int g
5	22	4.4	9858	8 ACC38255	ACC38255 Simocycli
6	22	4.4	37856	6 AAL11992	AAL11992 S. cellul
7	20	4.0	588	6 ABQ91233	ABQ91233 M. capsul
8	20	4.0	1091	3 AAF12516	AAF12516 Aspergill
9	20	4.0	5275	2 AAV26962	AAV26962 Bovine pa
10	20	4.0	5275	2 AAT95857	AAT95857 Bovine pa
11	20	4.0	5275	2 AAZ25053	AAZ25053 Bovine pa
12	20	4.0	5275	2 AAV82483	AAV82483 Bovine pa
13	20	4.0	5275	3 AAZ89296	AAZ89296 Bovine ca
14	20	4.0	5275	6 AAI72120	AAI72120 cDNA enco
15	19	3.8	558	9 ADC08850	ADC08850 Corn DNA
16	19	3.8	1172	2 AAV69527	AAV69527 Soil bact
17	19	3.8	1178	2 AAV69562	AAV69562 Soil deri
18	19	3.8	1818	7 ACC44559	ACC44559 Maize gra
19	19	3.8	1915	2 AAX60319	AAX60319 DNA seque
20	19	3.8	2223	7 ACC44560	ACC44560 Maize 797
21	19	3.8	2267	2 AAX63355	AAX63355 Granule b
22	19	3.8	3738	7 ACA25719	ACA25719 Prokaryot
23	19	3.8	4800	2 AAV29752	AAV29752 Zea may

c	24	19	3.8	4800	7 ABX09931	Abx09931 DNA encod
c	25	19	3.8	6029	4 AAK82712	AAK82712 Human imm
c	26	19	3.8	21185	3 AAG63350	AAG63350 Streptomy
c	27	19	3.8	24379	2 AAT93095	AAT93095 Streptomy
c	28	19	3.8	24379	2 AAV25925	AAV25925 Streptomy
c	29	19	3.8	63184	3 AAG63346	AAG63346 Streptomy
c	30	19	3.8	349980	6 ABQ81846	ABQ81846 Bifidobac
c	31	18	3.6	175	4 AAI29319	AAI29319 Colon tum
c	32	18	3.6	175	7 ABZ33505	ABZ33505 Human col
c	33	18	3.6	313	6 ABN18020	ABN18020 Human ORF
c	34	18	3.6	418	8 ACH19833	ACH19833 Human adu
c	35	18	3.6	457	5 AAH83034	AAH83034 Human ova
c	36	18	3.6	498	2 AAV29373	AAV29373 Human sma
c	37	18	3.6	537	5 AAS75981	AAS75981 DNA encod
c	38	18	3.6	622	4 AAL01672	AAL01672 Human rep
c	39	18	3.6	622	4 ABA07591	ABA07591 Human ova
c	40	18	3.6	633	5 AAS81267	AAS81267 DNA encod
c	41	18	3.6	819	8 ACC83519	ACC83519 Human exo
c	42	18	3.6	909	7 ACA44265	ACA44265 Prokaryot
c	43	18	3.6	1699	4 AAI14082	AAI14082 Human CDN
c	44	18	3.6	1703	6 ABL90390	ABL90390 Human pol
c	45	18	3.6	1715	2 AAZ24887	AAZ24887 Human sec

#### ALIGNMENTS

RESULT 1  
AAV21187  
ID AAV21187 standard; DNA; 53789 BP.  
XX  
AC AAV21187;  
XX  
DT 24-JUL-1998 (first entry)  
XX  
DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.  
XX  
KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
KW polyketide synthase; actinomycete; ansamycin; ds.  
XX  
OS Amycolatopsis mediterranei.  
XX  
FH Key Location/Qualifiers  
CDS 1825..15543  
FT /\*tag= a  
FT /label= ORF\_A  
FT /product= "Polyketide synthase"  
FT 15550..30759  
FT /\*tag= b  
FT /label= ORF\_B  
FT /product= "polyketide synthase"  
FT 30895..36060  
FT /\*tag= C  
FT /label= ORF\_C  
FT /product= "polyketide synthase"  
FT 36259..41325  
FT /\*tag= d  
FT /label= ORF\_D  
FT /product= "polyketide synthase"  
FT 41373..51614  
FT /\*tag= e  
FT /label= ORF\_E  
FT /product= "polyketide synthase"  
FT 51713..52993  
FT /\*tag= f  
FT /label= ORF\_F  
FT /product= "polyketide synthase"

WO9807868-A1.

26-FEB-1998.

18-AUG-1997; 97WO-EP004495.

XX 20-AUG-1996; 96EP-00810551.  
XX (NOVS ) NOVARTIS AG.  
XX Schupp T, Toupet C, Engel N;  
XX WPI; 1998-169172/15.  
DR P-PSDB; AAW52845, AAW52846, AAW52847, AAW52848, AAW52849, AAW52850.  
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used to  
XX produce rifamycin and rifamycin analogues.  
XX Claim 4; Page 53-102; 205pp; English.  
XX The present sequence represents a Amycolatopsis mediterranei rifamycin  
XX synthesis gene cluster DNA fragment from the present invention. The DNA  
XX fragment comprises a DNA region involved directly or indirectly in the  
XX gene cluster responsible for rifamycin synthesis, including the adjacent  
XX DNA regions to the right and left which, by reason of their function in  
XX connection with rifamycin biosynthesis, qualify as constituents of this  
XX rifamycin gene cluster, and functional fragments, derivatives or  
XX constituents of these. The Amycolatopsis mediterranei rifamycin synthesis  
XX gene cluster DNA fragment can be used for producing rifamycin, rifamycin  
XX analogues or precursors. It can also be used for inactivating or  
XX modifying genes involved in ansamycin or rifamycin biosynthesis. The DNA  
XX can be used for constructing mutant actinomycetes strains from which the  
XX natural rifamycin or ansamycin biosynthesis gene cluster has been partly  
XX or completely deleted. The DNA fragment can be used for assembling a  
XX library of polyketide synthases, which can be used for assembling a  
XX library of polyketides. A hybridisation probe of the invention can be  
XX used for identifying DNA fragments involved in the biosynthesis of  
XX ansamycins

SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 U; 0 Other;  
Query Match 100.0%; Score 501; DB 2; Length 53789;  
Best Local Similarity 100.0%; Pred. No. 7.1e-230;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTCTACAGTCGGGACGACGCGGCGGCCCAAGAGGCGTGTTCACGCGGCAAC 60  
DB 1825 ATGTTCTACAGTCGGGACGACGCGGCGGCCCAAGAGGCGTGTTCACGCGGCAAC 1884

QY 61 TGCCTCTGGTCCGCTTCTGTCTAGTCGCGCTCCCGGTTGTCGGACCGACCGG 120  
DB 1885 TGCCTCTGGTCCGCTTCTGTCTAGTCGCGCTCCCGGTTGTCGGACCGACCGG 1944

QY 121 GTGCTCTGGCGGCTCCCGCTGTTCACAGCGCTTTTCGACATCGGCTCGTCTCTCGGCC 180  
DB 1945 GTGCTCTGGCGGCTCCCGCTGTTCACAGCGCTTTTCGACATCGGCTCGTCTCTCGGCC 2004

QY 181 ACCGTGTCTGGGCGACGCTCGGATCGCGATCGCGACGCTCGCGGACGCGATCGG 240  
DB 2005 ACCGTGTCTGGGCGACGCTCGGATCGCGATCGCGACGCTCGCGGACGCGATCGG 2064

QY 241 CTGATCGAGCGGAGAGCTCGACCTTCTGCGCGGCTGCGACCTACCACTCG 300  
DB 2065 CTGATCGAGCGGAGAGCTCGACCTTCTGCGCGGCTGCGACCTACCACTCG 2124

QY 301 GTGCGGCGCGCGCGGCGGCTGCGGATTCGCGCGCGGAGCTCGGCTGCGCGG 360  
DB 2125 GTGCGGCGCGCGCGGCGGCTGCGGATTCGCGCGCGGAGCTCGGCTGCGCGG 2184

QY 361 GCGCGGCTCTCGGCGCGGCTGCGGATTCGAGAGCTTCGAGAGCTTCGCGGCTCGG 420  
DB 2185 GCGCGGCTCTCGGCGCGGCTGCGGATTCGAGAGCTTCGAGAGCTTCGCGGCTCGG 2244

QY 421 ATCGACCGCTACGCGACGACCTCGGCGGCTGCGGAGATCACCATGACCGCGGAGCG 480  
DB 2245 ATCGACCGCTACGCGACGACCTCGGCGGCTGCGGAGATCACCATGACCGCGGAGCG 2304

QY 481 GCCCGCGCTCGAGGCGCTCGTGC 501

DB 2305 GCCCGCGTCCGAGGGCTCGTGC 2325

RESULT 2  
AAL61170  
ID AAL61170 standard; DNA; 14055 BP.  
XX  
AC AAL61170;  
XX  
DT 22-SEP-2003 (first entry)  
XX  
DE Actinosynnema pretiosum polyketide synthase (PKS) gene #1.  
XX  
KW Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;  
XX gene; ds.  
XX  
OS Actinosynnema pretiosum.  
XX  
PN WO2003045312-A2.  
XX  
PD 05-JUN-2003.  
XX  
PF 21-NOV-2002; 2002WO-US037547.  
XX  
PR 21-NOV-2001; 2001US-0332158P.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Floss HG, Yu T, Leistner B;  
XX  
PP WPI; 2003-493374/46.  
XX  
PT Novel maytansinoid produced by bacterial host cell transformed with  
PT expression vector comprising open reading frame from ansamitocin gene  
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX  
PS Disclosure; Page 45-53; 160pp; English.  
XX  
CC The invention relates to maytansinoid produced by bacterial host cell  
CC transformed with expression vector comprising open reading frame from  
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
CC useful as a potent antitumour agent. The present sequence is A. pretiosum  
CC ansamitocin gene cluster I polyketide synthase (PKS) gene  
XX  
SQ Sequence 14055 BP; 1304 A; 5585 C; 5518 G; 1648 T; 0 U; 0 Other;

Query Match 5.2%; Score 26; DB 7; Length 14055;  
Best Local Similarity 100.0%; Pred. No. 0.012;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 ATCGAGCGCTTACGCGAGCAGCCGAGAC 446  
DB 904 ATCGAGCGCTTACGCGAGCAGCCGAGAC 929

RESULT 3  
AAL61224/c  
ID AAL61224 standard; DNA; 82745 BP.  
XX  
AC AAL61224;  
XX  
DT 22-SEP-2003 (first entry)  
XX  
DE Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.  
XX  
KW Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX  
OS Actinosynnema pretiosum.  
XX  
PN WO2003045312-A2.  
XX  
PD 05-JUN-2003.



```
XX 21-NOV-2002; 2002WO-US037547.
XX PF
XX PR
XX 21-NOV-2001; 2001US-0332158P.
XX PA
XX (UNIW ) UNIV WASHINGTON.
XX Floss HG, Yu T, Leistner E;
XX WPI; 2003-493374/46.
XX
XX Novel maytansinoid produced by bacterial host cell transformed with
XX expression vector comprising open reading frame from ansamitocin gene
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
XX Claim 7; Page 105-152; 160pp; English.
XX
XX The invention relates to maytansinoid produced by bacterial host cell
XX transformed with expression vector comprising open reading frame from
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
XX useful as a potent antitumor agent. The present sequence is A. pretiosum
XX ansamitocin ansamitocin biosynthetic gene cluster I
XX
XX Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;
XX
XX Query Match 5.2%; Score 26; DB 7; Length 82746;
XX Best Local Similarity 100.0%; Pred. No. 0.0095;
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 421 ATCGAGCGCTACGGCAGCAGCCGAGAC 446
Db 36138 ATCGAGCGCTACGGCAGCAGCCGAGAC 36113

RESULT 4
AAL40781
ID AAL40781 standard; DNA; 88421 BP.
AC AAL40781;
XX
XX 03-OCT-2002 (first entry)
XX
XX 88421nt genomic DNA of ramoplanin producing Actinoplanes sp.
XX
XX Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;
XX biosynthesis gene cluster; bioengineering; peptide synthetase module;
XX adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor;
XX chlorinate; lipdepsipeptide; gene; ds.
XX
XX Actinoplanes sp.
XX
XX Key Location/Qualifiers
XX CDS 2077..3078
XX /*tag= a
XX /*product= "Protein of ORF 1"
XX CDS 3118..4032
XX /*tag= b
XX /*product= "Protein of ORF 2"
XX CDS 4038..5048
XX /*tag= c
XX /*product= "Protein of ORF 3"
XX CDS complement(6665..5814)
XX /*tag= d
XX /*product= "Protein of ORF 4"
XX CDS complement(7703..6693)
XX /*tag= e
XX /*product= "Protein of ORF 5"
XX CDS complement(9464..8130)
XX /*tag= f
XX /*product= "Protein of ORF 6"
XX CDS 9691..10761
XX /*tag= g
XX /*product= "Protein of ORF 7"
XX
complement(12751..10829)
/*tag= h
/*product= "Protein of ORF 8"
complement(13617..12802)
/*tag= i
/*product= "Protein of ORF 9"
complement(15203..13614)
/*tag= j
/*product= "Protein of ORF 10"
15591..15863
/*tag= k
/*product= "Protein of ORF 11"
15880..19035
/*tag= l
/*product= "Protein of ORF 12"
19032..39713
/*tag= m
/*product= "Protein of ORF 13"
39713..65800
/*tag= n
/*product= "Protein of ORF 14"
65826..66530
/*tag= o
/*product= "Protein of ORF 15"
66546..67370
/*tag= p
/*product= "Protein of ORF 16"
67384..70059
/*tag= q
/*product= "Protein of ORF 17"
70099..70662
/*tag= r
/*product= "Protein of ORF 18"
70659..71906
/*tag= s
/*product= "Protein of ORF 19"
complement(73439..71964)
/*tag= t
/*product= "Protein of ORF 20"
complement(74216..73563)
/*tag= u
/*product= "Protein of ORF 21"
complement(75424..74213)
/*tag= v
/*product= "Protein of ORF 22"
75535..76464
/*tag= w
/*product= "Protein of ORF 23"
complement(78110..76449)
/*tag= x
/*product= "Protein of ORF 24"
complement(79864..78107)
/*tag= y
/*product= "Protein of ORF 25"
complement(81624..79861)
/*tag= z
/*product= "Protein of ORF 26"
complement(81909..81682)
/*tag= aa
/*product= "Protein of ORF 27"
complement(82346..82062)
/*tag= ab
/*product= "Protein of ORF 28"
82587..84446
/*tag= ac
/*product= "Protein of ORF 29"
84481..85548
/*tag= ad
/*product= "Protein of ORF 30"
85556..86845
/*tag= ae
/*product= "Protein of ORF 31"
87372..86803
```

```

FT      /tag= af
FT      /product= "Protein of ORF 32"
CDS     87494.88420
FT      /tag= ag
FT      /product= "Protein of ORF 33"

```

PN WO200231155-A2.

18-APR-2002.

15-OCT-2001: 2001WO-CA0001462

13-OCT-2000 2000US-0239924P

PR 12-APR-2001; 2001US-0283296F.  
PR 24-JUN-2001; 2001US-00810813

XX  
DA  
(PCOB-) ECOTR PROCTENCS TM

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

[illegible]

DR P-PSDB; AAO22146, AA

DR AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165  
DR AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,  
DR AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.

Novel isolated ramoplanin biosynthetic pathway polypeptide useful for chemically modifying biological molecule that is a substrate for a polypeptide encoded by a ramoplanin biosynthesis gene cluster.

PS Disclosure: page 87-135: 212pp: English.

The invention relates to an isolated ramoplanin biosynthetic pathway polypeptide selected from a polypeptide of open reading frames (ORF) 1-32. The isolated polypeptides are useful for chemically modifying a biological molecule that is a substrate for a polypeptide encoded by a ramoplanin biosynthesis gene cluster, by contacting the biological molecule with the isolated polypeptide, where the polypeptide chemically modifies the biological molecule. The method comprises contacting the biological molecule with at least two different polypeptides encoded by ramoplanin ORFs 1-31. The polypeptides are useful for directing the biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated gene cluster comprising the ORFs is useful as a substrate for bioengineering of antibiotic structures. An isolated polypeptide or its encoding nucleic acid sequence is useful for generating derivatives of ramoplanin, for improving production or for producing variants of other antibiotics of the peptide class. The isolated polypeptides are useful for synthesis of ramoplanin *in vivo* or *in vitro*, as an adenylation domain in conjunction with other peptide synthetase modules and allowing the incorporation of Thr into a peptide antibiotic precursor, for modifying fatty acid structure and/or enhancing fatty acid incorporation into the peptide antibiotic structure, for production of an hydroxyphenylglycine (HPG)-containing peptide antibiotic, for enhancing secretion of ramoplanin or its variants and derivatives, for enhancing uptake of precursors for ramoplanin biosynthesis, for enhancing production of ramoplanin products or its variants or derivatives, to chlorinate HPG of a peptide antibiotic precursor, and for designing specific nucleotide probes and primers for identifying and isolating putative lipopeptide-producing microorganisms. This polynucleotide sequence represents the 8842nt genomic DNA of a ramoplanin producing *Actinoplanes* sp. microorganism of the invention.

SQ Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 U; 0 Other;  
Query Match 5.2%; Score 26; DB 6; Length 89421;  
Best Local Similarity 100.0%; Pred.No. 0.0094;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QV 20 CGACCGGGCGGCCCAAGGGCGTCGTG 45

38404 CGACCGGGGGCCCAAGGGCGTCTG 38429

DR WPI; 2003-268202/26.  
DR P-PSDB; ABR42548, ABR42549, ABR42550, ABR42551, ABR42552, ABR42553,  
DR ABR42554, ABR42555.  
XX  
XX Novel polypeptide encoded by open reading frame of aminocoumarin  
PT biosynthetic gene cluster, useful for synthesizing aminocoumarins useful  
PT for treating bacterial infections and malignancies.  
XX  
PS Claim 2; Fig 3; 124pp; English.  
XX  
XX The present sequence is a partial sequence of the sinocyclinone  
CC biosynthetic gene cluster of Streptomyces antibioticus. It includes open  
CC reading frames (ORFs) that show sequence similarity to ORFs of the  
CC novobiocin biosynthetic gene cluster. The invention provides  
CC aminocoumarin (e.g. sinocyclinone) biosynthetic nucleic acids and  
CC polypeptides, a method for producing hybrid antibiotics using a  
CC combination of various regions of biosynthetic gene clusters, and a  
CC pharmaceutical composition containing an aminocoumarin compound for use  
CC in the treatment of Gram-positive bacterial infections and of malignant  
CC diseases  
XX  
SQ Sequence 9858 BP; 1512 A; 3292 C; 3482 G; 1572 T; 0 U; 0 Other;  
Query Match 4.4%; Score 22; DB 8; Length 9858;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 20 CGACCGGGCGGCCCAAGGGCGT 41  
Db 4134 CGACCGGGCGGCCCAAGGGCGT 4155  
RESULT 6  
AA11992  
ID AA11992 standard; DNA; 37856 BP.  
AC AA11992;  
XX  
XX 15-SEP-2003 (revised)  
DT 07-AUG-2000 (first entry)  
XX  
DE S. cellulose DNA encoding polyketide and heteropolyketide enzymes.  
XX  
XX Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;  
KW epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;  
KW plant-protection; ds.  
XX  
OS Polyangium cellulosum.  
XX  
FH Location/Qualifiers  
FT complement(3398..6100)  
FT /\*tag= a  
FT /product= "ORF1-tRNA synthetase"  
FT /note= "gtg start codon"  
FT complement(6374..7111)  
FT /\*tag= b  
FT /product= "ORF2-monoxygenase"  
FT complement(8433..9550)  
FT /\*tag= c  
FT /product= "ORF3-aminotransferase"  
FT /note= "AGT start codon given in the specification"  
FT 9855..11393  
FT /\*tag= d  
FT /product= "ORF4- tyrosine/DOPA-Decarboxylase"  
FT /note= "GTG start codon"  
FT 12212..13658  
FT /\*tag= e  
FT /product= "ORF5-3-oxoacyl-ACP-reductase"  
FT /note= "ACC start codon"  
FT 15374..19984  
FT /\*tag= f  
FT /product= "ORF6-polyketide synthase"  
FT 20003..27889

FT /\*tag= g  
FT /product= "ORF7-peptide synthetase"  
FT 28251..29400  
FT /\*tag= h  
FT /product= "ORF8-transpeptidase"  
FT complement(30040..31720)  
FT /\*tag= i  
FT /product= "ORF9-regulation element"  
FT /note= "CGC stop codon"  
FT 31982..32932  
FT /\*tag= j  
FT /product= "ORF10-transcription regulator"  
FT 33128..33613  
FT /\*tag= k  
FT /product= "ORF11-regulation element"  
FT /note= "GTG start codon"  
FT 33661..34077  
FT /\*tag= l  
FT /product= "ORF12-regulation element"  
FT complement(35255..35616)  
FT /\*tag= m  
FT /product= "ORF13-transcription regulator"  
FT complement(35730..36242)  
FT /\*tag= n  
FT /product= "ORF14-transcription regulator"  
FT /note= "GTG start codon"  
XX  
XX DE19846493-A1.  
XX  
XX 13-APR-2000.  
XX  
XX 09-OCT-1998; 98DE-01046493.  
XX  
XX 09-OCT-1998; 98DE-01046493.  
XX (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.  
XX Beyer S, Mueller R;  
XX WPI; 2000-294101/26.  
XX  
XX DNA sequence coding for products involved in the biosynthesis of  
XX polyketide or heteropolyketide compounds, especially epothilone.  
XX  
XX Claim 3; Page 20-33; 36pp; German.  
XX  
XX This invention describes a novel DNA sequence (I) whose expression  
XX products effect or are involved in the enzymatic biosynthesis,  
XX mutasynthesis or partial synthesis of polyketide or heteropolyketide  
XX compounds (II). (I) can be inserted into an expression vector and used to  
XX transform or transfect prokaryotic or eukaryotic cells with the aim of  
XX obtaining strains that produce large amounts of polyketide or  
XX heteropolyketide compounds, especially epothilones, which have cytotoxic  
XX and/or immunosuppressant and antibiotic and antifungal activities and are  
XX useful as plant-protection agents. This sequence represents the DNA  
XX sequence isolated from Sorangium cellulosum which is described in the  
XX method of the invention. (Updated on 15-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 37856 BP; 5655 A; 13666 C; 12913 G; 5622 T; 0 U; 0 Other;  
Query Match 4.4%; Score 22; DB 3; Length 37856;  
Best Local Similarity 100.0%; Pred. No. 0.85;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 20 CGACCGGGCGGCCCAAGGGCGT 41  
Db 21906 CGACCGGGCGGCCCAAGGGCGT 21927  
RESULT 7  
ABQ91233  
ID ABQ91233 standard; DNA; 588 BP.  
XX

AC ABQ91233;  
 XX  
 DT 01-OCT-2002 (first entry)  
 XX  
 DE M. capsulatus gene #1218 for DNA array.  
 XX  
 KW Micro array; gene; ds; differential expression; gene expression.  
 XX  
 OS Methylococcus capsulatus.  
 XX  
 PN WO200255655-A2.  
 XX  
 PD 18-JUL-2002.  
 XX  
 PF 14-JAN-2002; 2002WO-N0000019.  
 XX  
 PR 12-JAN-2001; 2001NO-00000235.  
 XX  
 PR 12-JAN-2001; 2001NO-00000239.  
 XX  
 PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.  
 PA (TIGR-) TIGR.  
 XX  
 PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;  
 PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;  
 PI Salzberg SL;  
 XX  
 DR WPI; 2002-557818/59.  
 XX  
 XX Novel DNA array useful for determining differential expression of  
 PT Methylococcus capsulatus genes, comprises polynucleotides or  
 PT oligonucleotides representative for a selective number of Methylococcus  
 PT capsulatus genes.  
 XX  
 PS Claim 14; Page 502; 678pp; English.  
 XX  
 CC The invention relates to a novel DNA array giving a representation of a  
 CC number of Methylococcus capsulatus genes. The method of the invention is  
 CC useful for determination of the differential expression of the genes of  
 CC M. capsulatus, and for studying gene expression on a genomic scale and in  
 CC gene expression assays of M. capsulatus genes. The sequences shown in  
 CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the  
 CC invention  
 XX  
 SQ Sequence 588 BP; 105 A; 196 C; 196 G; 91 T; 0 U; 0 Other;  
 Query Match 4.0%; Score 20; DB 6; Length 588;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 273 CGGCGTGCCGACCACTACC 292  
 DB 447 CGGCGTGCCGACCACTACC 466  
 RESULT 8  
 AAF12516/c  
 ID AAF12516 standard; cDNA; 1091 BP.  
 XX  
 AC AAF12516;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Aspergillus oryzae EST SEQ ID NO:5039.  
 XX  
 KW Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Aspergillus oryzae.  
 XX  
 PN WO200056762-A2.  
 XX  
 XX 28-SEP-2000.  
 XX  
 XX 22-MAR-2000; 2000WO-US007781.  
 XX  
 XX 22-MAR-1999; 99US-00273623.  
 XX  
 PA (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 XX  
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 PI WPI; 2000-594572/56.  
 XX  
 DR Monitoring differential expression of genes in filamentous fungal cells  
 XX uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags.  
 XX  
 PS Claim 88; Page 2116; 3161pp; English.  
 XX  
 CC The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring the  
 CC global expression of genes from FF cells allows the production potential  
 CC of the microorganisms to be improved. New genes may be discovered.  
 CC possible functions of unknown open reading frames can be identified and  
 CC gene copy number variation and stability can be monitored. The expression  
 CC of genes can be used to study how FF cells adapt to changes in culture  
 CC conditions, environmental stress, spore morphogenesis, recombination,  
 CC metabolic or catabolic pathway engineering. Using ESTs provides several  
 CC advantages over genomic or random cDNA clones including elimination of  
 CC redundancy as one spot on an array equals one gene or open reading frame,  
 CC and organisation of the microarrays based on function of the gene  
 CC products to facilitate analysis of the results. AAF07478 to AAF11247  
 CC represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents  
 CC ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from  
 CC Aspergillus oryzae; and AAF14879 to AAF15337 represents ESTs from  
 CC Trichoderma reesei, which are all specifically claimed in the present  
 CC invention  
 XX  
 SQ Sequence 1091 BP; 240 A; 305 C; 308 G; 238 T; 0 U; 0 Other;  
 Query Match 4.0%; Score 20; DB 3; Length 1091;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 308 CGGCCGCGCAGCGCGTTTC 327  
 DB 656 CGGCCGCGCAGCGCGTTTC 637  
 RESULT 9  
 AAV26962  
 ID AAV26962 standard; cDNA to mRNA; 5275 BP.  
 XX  
 AC AAV26962;  
 XX  
 DT 01-SEP-1998 (first entry)  
 XX  
 DE Bovine parathyroid calcium receptor 1 gene 5kb fragment.  
 XX  
 KW ss; calcium ion concentration; parathyroid hormone; homeostasis; kidney;  
 KW calcium receptor; detection.  
 XX  
 OS Bos sp.  
 XX  
 XX Key Location/Qualifiers  
 FH 515. .3772  
 FT CDS



```

XX 05-OCT-1999.
XX 03-OCT-1997; 97US-00943986.
XX 23-FEB-1993; 93US-00009389.
XX 22-OCT-1993; 93US-00141248.
XX 19-AUG-1994; 94US-00292827.
XX 21-OCT-1994; 94WO-US012117.
XX 08-DEC-1994; 94US-00353784.
XX 07-JUN-1995; 95US-00484565.
XX (NPSP-) NPS PHARM INC.
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX Brown EM, Hebert SC, Garrett JE;
XX WPI; 1999-571274/48.
XX P-PSDB; AAY41778.
XX Nucleic acids encoding protein calcium receptors useful for identifying
XX modulators of receptor expression and activity and for the production of
XX antigens specific for calcium receptors.
XX Claim 5; Fig 47; 174pp; English.
XX The present sequence encodes bovine parathyroid calcium receptor 1
XX (BoPCar 1). Calcium receptor polynucleotides may be used: (i) for
XX producing receptor proteins (or fragments) useful for determining
XX structure and activity relationships, for assaying molecular activity on
XX the receptor (i.e. identifying modulators of receptor function) and for
XX producing antibodies specific for the receptor; (ii) for sequencing the
XX normal form of the nucleic acids (the derived sequence may be compared
XX with other receptors to identify conserved sequences, mutations and
XX variations that may influence calcium receptor activity and to determine
XX target sites for antisense molecules, ribozymes, hybridization probes and
XX polymerase chain reaction (PCR) amplification primers; (iii) as
XX hybridization probes to detect the presence of similar sequences in
XX samples; and (iv) as PCR primers to generate particular nucleic acid
XX sequence regions; (e.g. to generate regions to be probed by hybridization
XX detection probes)
XX Sequence 5275 BP; 1277 A; 1475 C; 1316 G; 1207 T; 0 U; 0 Other;
XX Query Match 4.0%; Score 20; DB 2; Length 5275;
XX Best Local Similarity 100.0%; Pred. No. 9.8;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 362 GCGGGTCTCTCGCGCGGG 381
DB 223 GCGGGTCTCTCGCGCGGG 242
RESULT 12
AAV82483
ID AAV82483 standard; cDNA to mRNA; 5275 BP.
XX AC AAV82483;
XX 19-MAR-1999 (first entry)
XX Bovine parathyroid calcium receptor BoPCar 1 encoding cDNA.
XX Parathyroid calcium receptor; inorganic ion receptor; osteoporosis;
XX calcium homeostasis; hyperparathyroidism; seizure; stroke; epilepsy;
XX spinal cord injury; hypoxia-induced nerve cell damage; cardiac arrest;
XX neonatal distress; neurodegenerative disease; Alzheimer's disease;
XX Huntington's disease; Parkinson's disease; dementia; muscle tension;
XX depression; anxiety; ss.
XX Bos sp.
XX Key Location/Qualifiers

```

```

FT CDS 515..3772
FT FT /*tag= a
XX PN US5858684-A.
XX 12-JAN-1999.
XX 07-JUN-1995; 95US-00480751.
XX 23-AUG-1991; 91US-00749451.
XX 11-FEB-1992; 92US-00834044.
XX 21-AUG-1992; 92US-00934161.
XX 12-FEB-1993; 93US-00017127.
XX 23-FEB-1993; 93US-00009389.
XX 22-OCT-1993; 93US-00141248.
XX 19-AUG-1994; 94US-00292827.
XX 21-OCT-1994; 94WO-US012117.
XX 08-DEC-1994; 94US-00353784.
XX (NPSP-) NPS PHARM INC.
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX Del Mar EG, Balandrin MF, Van Wagenen BC, Nemeth EF, Brown EM;
XX Garrett JE, Hebert SC;
XX WPI; 1999-119871/10.
XX P-PSDB; AAW89563.
XX Screening for calcium receptor-active compounds - by recombinant
XX expression of nucleic acid encoding calcium receptor and determining the
XX effect of compounds on calcium receptor activity.
XX Claim 1; Fig 47; 176pp; English.
XX A method has been developed of screening for a compound able to affect
XX one or more activities of a calcium receptor (CR) comprises: (A)
XX contacting a recombinant cell with a test compound, where the recombinant
XX cell comprises a recombinant nucleic acid expressing the CR, provided
XX that the cell does not have functional CR expression from endogenous
XX nucleic acid; (B) determining the ability of the test compound to affect
XX one or more activities of the calcium receptor; and (C) comparing the
XX CR activities in a cell not comprising the recombinant nucleic acid. The
XX present sequence encodes bovine parathyroid CR, designated a BoPCar 1.
XX The nucleic acid sequence of BoPCar 1 can be used as part of the
XX recombinant nucleic acid in the method described above. The compounds
XX identified can be used to treat diseases or disorders characterised by
XX abnormal calcium homeostasis, e.g. hyperparathyroidism, osteoporosis and
XX other bone and mineral-related disorders. They can also be used for the
XX treatment of diseases and disorders associated with disrupted Ca2+
XX responses, e.g. seizures, stroke, spinal cord injury, hypoxia-induced
XX nerve cell damage such as in cardiac arrest or neonatal distress,
XX epilepsy, neurodegenerative diseases such as Alzheimer's disease,
XX Huntington's disease and Parkinson's disease, dementia, muscle tension,
XX depression, and anxiety
XX Sequence 5275 BP; 1277 A; 1475 C; 1316 G; 1207 T; 0 U; 0 Other;
XX Query Match 4.0%; Score 20; DB 2; Length 5275;
XX Best Local Similarity 100.0%; Pred. No. 9.8;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 362 GCGGGTCTCTCGCGCGGG 381
DB 223 GCGGGTCTCTCGCGCGGG 242
RESULT 13
AAZ89296
ID AAZ89296 standard; cDNA to mRNA; 5275 BP.
XX AC AAZ89296;
XX AAZ89296;

```

DT 09-JUN-2000 (first entry)

DE Bovine calcium receptor BoPCar1 cDNA.

XX Calcium receptor; treatment; calcimimetic; calcilytic; osteopathic;

KW cerebroprotective; cytosolic; neuroprotective; dermatological;

KW tranquilizer; vulnerar; antitumor; immunosuppressive; hypotensive;

KW cardiac; parathyroid hormone; osteoporosis; calcitonin secretion;

KW hyperparathyroidism; Paget's disease; bovine; ss.

XX Bos taurus.

XX Key Location/Qualifiers

FT 515..3772

FT CDS /\*tag= a

FT /product= "calcium receptor"

XX US6031003-A.

XX 29-FEB-2000.

XX 07-JUN-1995; 95US-00484719.

XX 23-AUG-1991; 91US-00749451.

XX 11-FEB-1992; 92US-00834044.

XX 21-AUG-1992; 92US-00934161.

XX 12-FEB-1993; 93US-00017127.

XX 23-FEB-1993; 93US-00009389.

XX 22-OCT-1993; 93US-00141248.

XX 19-AUG-1994; 94US-00292827.

XX 21-OCT-1994; 94WO-US012117.

XX 08-DEC-1994; 94US-00353784.

XX (NPSP-) NPS PHARM INC.

PA (BGHM) BRIGHAM & WOMENS HOSPITAL.

XX Balandrin MF, Delmar EG, Moe ST, Nemeth EF, Van Wagenen BC;

XX WPI; 2000-301969/26.

DR P-PSDB; AAB51825.

XX Treating disorders or diseases in a patient by modulating inorganic ion

PT receptor activities especially calcium receptor for hyperparathyroidism,

PT by administering a calcimimetic or calcilytic compound.

XX Example 25; Col 107-116; 194pp; English.

XX This invention describes a novel method of treating disorders by

CC modulating calcium receptor activity in vitro by administering a

CC calcimimetic, or a calcilytic compound. The products of the invention

CC have osteopathic, cerebroprotective, cytosolic, neuroprotective,

CC dermatological, tranquilizer, vulnerar, antitumor, immunosuppressive,

CC hypotensive and cardiac activity. The method is suitable for reducing

CC parathyroid hormone level in a patient to that of a normal individual,

CC treating a patient having osteoporosis, to inhibit bone resorption, and

CC to stimulate calcitonin secretion in vitro or in vivo. The level of

CC parathyroid hormone is reduced to cause a decrease in plasma Ca<sup>2+</sup>. The

CC method is useful in treating disorders in humans such as

CC hyperparathyroidism, Paget's disease and osteoporosis. Also for treatment

CC or prevention, based on the affected cells, of other disorders and

CC conditions like seizures, stroke, head trauma, spinal cord injury,

CC hypoxia-induced nerve cell damage such as in cardiac arrest or neonatal

CC distress, epilepsy, Alzheimer's disease, Huntington's disease, panic

CC Parkinson's disease, dementia, muscle tension, depression, anxiety, panic

CC disorder, OCD (not defined), post-traumatic stress disorder,

CC schizophrenia, neuroleptic malignant syndrome and Tourette's syndrome,

CC diseases involving excess water reabsorption by the kidney such as

CC syndrome of inappropriate ADH secretion (SIADH), cirrhosis, congestive

CC heart failure and nephrosis, hypertension, preventing and/or decreasing

CC renal toxicity from cationic antibiotics (e.g. aminoglycoside

CC antibiotics), gut motility disorders such as diarrhoea and spastic colon,

CC GI (gastrointestinal) ulcer diseases, GI diseases with excessive calcium

CC absorption such as sarcoidosis and autoimmune diseases and organ

CC CC transplanted rejection. This sequence encodes the bovine calcium receptor

CC BoPCar1 which is described in the method of the invention

XX

SQ Sequence 5275 BP; 1277 A; 1475 C; 1316 G; 1207 T; 0 U; 0 Other;

Query Match 4.0%; Score 20; DB 3; Length 5275;

Best Local Similarity 100.0%; Pred. No. 9.8;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGTCTCGCGCGCGG 381

DB 223 GCGCGTCTCGCGCGCGG 242

RESULT 14

AAI72120

ID AAI72120 standard; cDNA to mRNA; 5275 BP.

XX

AC AAI72120;

XX

DT 07-AUG-2003 (revised)

DT 25-MAR-2002 (first entry)

XX

DE cDNA encoding BoPCar1.

XX Calcium receptor; bovine; human; parathyroid; calcium receptor; thyroid;

XX C-cell; inorganic ion receptor; homeostasis; sarcoidosis;

XX hyperparathyroidism; osteoporosis; central nervous system; seizure;

XX stroke; head trauma; spinal cord injury; organ transplant rejection;

XX hypoxia-induced nerve cell damage; cardiac arrest; neonatal distress;

XX epilepsy; neurodegenerative disease; Alzheimer's disease; cirrhosis;

XX Huntington's disease; Parkinson's disease; dementia; depression; anxiety;

XX panic disorder; obsessive-compulsive disorder; spastic colon;

XX post-traumatic stress disorder; schizophrenia; diarrhoea; kidney;

XX neuroleptic malignant syndrome; Tourette's syndrome; gut motility;

XX inappropriate ADH secretion; SIADH; gastrointestinal ulcer disease;

XX congestive heart failure; nephrosis; hypertension;

XX aminoglycoside antibiotic; ss.

XX Bos taurus.

XX

XX Key Location/Qualifiers

FT 515..3772

FT CDS /\*tag= a

FT /product= "BoPCar1"

XX

PN US6313146-B1.

XX

XX 06-NOV-2001.

XX

XX 07-JUN-1995; 95US-00484159.

XX

XX 23-AUG-1991; 91US-00749451.

XX 11-FEB-1992; 92US-00834044.

XX 21-AUG-1992; 92US-00934161.

XX 12-FEB-1993; 93US-00017127.

XX 23-FEB-1993; 93US-00009389.

XX 22-OCT-1993; 93US-00141248.

XX 19-AUG-1994; 94US-00292827.

XX 21-OCT-1994; 94WO-US012117.

XX 08-DEC-1994; 94US-00353784.

XX (NPSP-) NPS PHARM INC.

XX Van Wagenen BC, Balandrin MF, Delmar EG, Nemeth EF;

XX WPI; 2002-081872/11.

XX P-PSDB; AAB47820.

XX Novel inorganic ion receptor-modulating compounds, useful for treating

PT e.g. hyperparathyroidism, osteoporosis, stroke, epilepsy, Alzheimer's

PT disease, dementia, depression, anxiety, hypertension, cirrhosis and

PT spastic colon.

XX Example 25; Fig 47; 227pp; English.

XX The sequences given in AAT72120-23 encode various calcium receptor

XX proteins. This sequence encodes bovine parathyroid calcium receptor. The

XX calcium receptor proteins are used, in conjunction with the compounds of

XX the invention, for structure determination, to assay a molecule's activity

XX on a receptor and to obtain antibodies to that receptor. The compounds of

XX the invention, which modulate inorganic ion receptors are useful for

XX treating and diagnosing diseases or disorders which can be treated by

XX modulating inorganic ion receptor activity. This is preferably a disease

XX or disorder characterized by abnormal inorganic ion homeostasis,

XX preferably abnormal calcium homeostasis (hyperparathyroidism,

XX osteoporosis and other bone and mineral-related disorders), an abnormal

XX level of a messenger whose production or secretion is affected by

XX inorganic ion receptor activity, and an abnormal level or activity of a

XX messenger whose function is affected by inorganic ion receptor activity.

XX These compounds are also useful for treating and diagnosing diseases or

XX disorders of the central nervous system such as seizures, stroke, head

XX trauma, spinal cord injury, hypoxia-induced nerve cell damage such as in

XX cardiac arrest or neonatal distress, epilepsy, neurodegenerative diseases

XX such as Alzheimer's disease, Huntington's disease and Parkinson's

XX disease, dementia, depression, anxiety, panic disorder, obsessive-

XX compulsive disorder, post-traumatic stress disorder, schizophrenia,

XX neuroleptic malignant syndrome and Tourette's syndrome, diseases

XX involving excess water reabsorption by the kidney such as inappropriate

XX ADH secretion (SIADH), cirrhosis, congestive heart failure, nephrosis,

XX hypertension, for preventing and/or decreasing renal toxicity from

XX cationic antibiotics (e.g. aminoglycoside antibiotics), gut motility

XX disorders such as diarrhoea, and spastic colon, gastrointestinal (GI)

XX ulcer diseases, GI diseases with excessive calcium absorption such as

XX sarcoidosis, and autoimmune diseases and organ transplant rejection.

XX (Updated on 07-AUG-2003 to correct OS field.)

SQ Sequence 5275 BP; 1277 A; 1475 C; 1316 G; 1207 T; 0 U; 0 Other;

Query Match 4.0%; Score 20; DB 6; Length 5275;

Best Local Similarity 100.0%; Pred. No. 9.8;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGCGCGCGG 381

DB 223 GCGCGGTCTCGCGCGCGG 242

RESULT 15

ADC08850/c

ID ADC08850 standard; DNA; 558 BP.

AC ADC08850;

DT 18-DEC-2003 (first entry)

XX Corn DNA sequence Seq ID1155 related to grain filling.

XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;

XX carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;

XX tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;

XX wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;

XX gene; ds; plant.

XX Zea mays.

XX WO200300905-A2.

PN 03-JAN-2003.

XX 21-JUN-2002; 2002WO-IB002450.

PF 22-JUN-2001; 2001US-0300112P.

PR 26-SEP-2001; 2001US-0325277P.

PR 20-DEC-2001; 2001US-0342327P.

XX

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;

PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;

XX WPI; 2003-229341/22.

XX New plant genes encoding polypeptides having an activity involved in or

XX associated with the synthesis, metabolism or degradation of carbohydrates

XX in the plant grain useful in generating plants having improved

XX nutritional properties.

PS Disclosure; SEQ ID NO 1155; 130pp; English.

XX This invention, in the area of plant biotechnology, relates to novel

XX polynucleotides comprising a nucleotide sequence encoding a protein which

XX is involved in or associated with the synthesis, metabolism or

XX degradation of carbohydrates in the plant grain and the expression of

XX which is up-regulated during grain filling. The plant is selected from

XX corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,

XX sugarbeet, wheat, and rice. The invention may be useful for the

XX improvement of protein, oil, starch, fibre and moisture content of the

XX cereal grains. In addition, carbohydrate levels may be modified to a more

XX desirable level using the present invention. The present sequence is a

XX DNA sequence from corn which showed homology to rice "grain filling"

XX genes of the invention. Note: The sequence data for this patent did not

XX form part of the printed specification, but was obtained in electronic

XX format directly from WIPO at ftp.wipo.int/pub/publishedpat\_sequences.

SQ Sequence 558 BP; 108 A; 173 C; 198 G; 79 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 9; Length 558;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCGCGGTCTCGCGCGG 289

DB 395 GCGCGGTCTCGCGCGG 377

RESULT 16

AAV69527/c

ID AAV69527 standard; DNA; 1172 BP.

AC AAV69527;

XX 15-MAR-1999 (first entry)

XX Soil bacteria peptide synthase clone ps32 DNA.

XX Peptide synthase; soil; lichen; antibiotic biosynthesis; humus;

XX therapeutic; immunosuppressor; antitumour agent; pathogen;

XX genetic diversity; ss.

XX Bacteria.

XX WO9853057-A2.

XX 26-NOV-1998.

XX 21-MAY-1998; 98WO-CA000488.

XX 22-MAY-1997; 97US-00861774.

XX (TERR-) TERRAGEN DIVERSITY INC.

PA Waters B, Miao VEW, Yap WH, Seow KT;

PI WPI; 1999-070158/06.

DR P-FSDB; AAW82675.

XX New degenerate primers - used for recovering antibiotic biosynthetic DNA

PT from soil/lichen material.



XX PS Claim 17; Page 25; 98pp; English.

XX CC This sequence encodes a peptide synthase from soil/lichen material. The

XX CC encoded protein is used in a method for the recovery of antibiotic

XX CC biosynthetic DNA from humic materials or lichen. The PCR products of the

XX CC invention have the potential to be used as therapeutic molecules

XX CC including antibiotics, immunosuppressors and antitumour agents. The

XX CC method allows access to the reservoir of genetic diversity in soil

XX CC pathogenic micro-organisms, in order to find new antibiotics. It also

XX CC allows access to novel biosynthetic genes/enzymes that can be used to

XX CC produce antibiotics or produce specific compounds, enzymatically, in

XX CC vitro

XX SQ Sequence 1172 BP; 209 A; 388 C; 398 G; 177 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 1172;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCTACACGTCGGCAGCAGC 23

|||||

Db 1171 TCTACACGTCGGCAGCAGC 1153

RESULT 17

AAV69562/c

ID AAV69562 standard; DNA; 1178 BP.

XX AC AAV69562;

XX XX

DT 17-OCT-2003 (revised)

DT 15-MAR-1999 (first entry)

XX XX

DE Soil derived peptide synthase clone ps25 DNA.

XX XX

KW Peptide synthase; soil; lichen; antibiotic biosynthesis; humus;

KW therapeutic; immunosuppressor; antitumour agent; pathogen;

KW genetic diversity; ss.

XX unidentifed.

OS

PN WC9853097-A2.

XX WC9853097-A2.

XX 26-NOV-1998.

XX 21-MAY-1998; 98WO-CA000488.

XX 22-MAY-1997; 97US-00861774.

XX (TERR-) TERRAGEN DIVERSITY INC.

PA Waters B, Miao VFW, Yap WH, Seow XT;

PI

DR WPI; 1999-070158/06.

DR P-PSDB; AAW62714.

XX

XX New degenerate primers - used for recovering antibiotic biosynthetic DNA

XX from soil/lichen material.

XX

XX Claim 17; Page 89-90; 98pp; English.

XX PS

XX CC This sequence encodes a peptide synthase clone, isolated from soil. This

XX CC protein is used in a method for the recovery of antibiotic biosynthetic

XX CC DNA from humic materials or lichen. The PCR products of the invention

XX CC have the potential to be used as therapeutic molecules including

XX CC antibiotics, immunosuppressors and antitumour agents. The method allows

XX CC access to the reservoir of genetic diversity in soil pathogenic micro-

XX CC organisms, in order to find new antibiotics. It also allows access to

XX CC novel biosynthetic genes/enzymes that can be used to produce antibiotics

XX CC or produce specific compounds, enzymatically, in vitro. (Updated on 17-

XX CC OCT-2003 to standardise OS field)

SQ Sequence 1178 BP; 184 A; 407 C; 397 G; 190 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 1178;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCTACACGTCGGCAGCAGC 23

|||||

Db 1177 TCTACACGTCGGCAGCAGC 1159

RESULT 18

ACC44559/c

ID ACC44559 standard; cDNA; 1818 BP.

XX AC ACC44559;

XX XX

DT 02-JUN-2003 (first entry)

XX XX

DE Maize granule-bound starch synthase encoding cDNA SEQ ID NO:7.

XX XX

KW Self-processing plant; plant; processing enzyme; alpha-amylase; grain;

KW pullulanase; alpha-glucosidase; glucose isomerase; glucoamylase;

KW mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;

KW maltodextrin; ethanol; fermentation; beverage; enzyme; gene; ss.

XX Zea mays.

XX

FH Key Location/Qualifiers

CDS 1..1818

FT /\*tag= a

FT /partial

FT /product= "granule-bound starch synthase (waxy)"

FT /note= "no stop codon given"

XX XX

PN WO2003018766-A2.

XX XX

XX 06-MAR-2003.

XX XX

PF 27-AUG-2002; 2002WO-US027129.

XX XX

PR 27-AUG-2001; 2001US-0315281P.

XX XX

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX XX

PI Lanahan MB, Basu SS, Batie CJ, Chen W, Craig J, Kinkema M;

XX XX

DR WPI; 2003-268420/26.

DR P-PSDB; ABP96605.

XX XX

PT Novel polynucleotide encoding hyperthermophilic processing enzymes e.g.

PT alpha-amylase, useful for producing plant to produce food products having

PT improved taste or fermentable substrates for ethanol.

XX XX

PS Example 2; Page 90-91; 158pp; English.

XX XX

XX CC The present invention describes polynucleotides which encode processing

XX CC enzymes (e.g. alpha-amylase, pullulanase, alpha-glucosidase, glucose

XX CC isomerase, or glucoamylase) that are optimised for expression in plants.

XX CC The polynucleotides encode mesophilic, thermophilic or hyperthermophilic

XX CC processing enzymes, which are activated under suitable conditions to act

XX CC upon the desired substrate. Also described are self-processing transgenic

XX CC plants and plant parts, e.g. grain, which express one or more of these

XX CC enzymes and have an altered composition that facilitates plant and grain

XX CC processing. Also described is a method (M) for converting starch to

XX CC starch-derived products in a transformed plant part (TPP), by activating

XX CC the starch processing enzyme contained in it. Transgenic grain is useful

XX CC for preparing maltodextrin. A transformed plant (TP) can be used to

XX CC produce food products having improved taste and to produce fermentable

XX CC substrates for ethanol and fermented beverages. (M) eliminates the need

XX CC to mill or physically disrupt the integrity of plant parts prior to

XX CC recovery of starch-derived products. The present sequence encodes maize

XX CC granule-bound starch synthase (waxy), which is given in the

CC exemplification of the present invention

XX SQ Sequence 1818 BP; 346 A; 596 C; 609 G; 267 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 7; Length 1818;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGCGGTGCCGACCACT 289

Db 1655 GCCGCGGTGCCGACCACT 1637

RESULT 19

AAX60319/c

ID AAX60319 standard; DNA; 1915 BP.

XX AC AAX60319;

XX DT 23-AUG-1999 (first entry)

XX DE DNA sequence of the maize waxy gene.

XX KW Non-glycogen-like polysaccharide production; fermentation; waxy gene;

XX KW starch synthesis enzyme; glycogen-synthesis enzyme; glycogen synthesis;

XX KW non-starch branching gene; amylopectin; amylose; plant-like starch; ss.

XX OS Zea mays.

XX WO9844780-A1.

XX PD 15-OCT-1998.

XX PF 03-APR-1998; 98WO-US006660.

XX PR 04-APR-1997; 97US-0042939P.

XX PA (EXSE-) EXSEED GENETICS LLC.

XX PI Guan H, Keeling PL;

XX WPI; 1998-568285/48.

XX DR P-PSDB; AAY16604.

XX PT Producing non-glycogen-like polysaccharides in bacteria, fungi or plants

XX PT - transformed with genes for enzymes involved in starch or glycogen

XX PT synthesis allows fermentative production of starches with engineered

XX PS properties.

XX Dislosure; Fig 49; 150pp; English.

XX The specification describes a method for the production of non-glycogen-

XX like polysaccharides in a host. The method comprises transforming a host,

XX suitable for fermentation, with genes encoding starch- or glycogen-

XX synthesis enzymes, and fermenting the transformants. The specification

XX also describes hosts transformed with a gene active in glycogen synthesis

XX and at least one non-starch branching gene, involved in production of

XX amylopectin or amylose in its original host. The method is used to

XX produce plant-like starches by fermentation and new starches in plants.

XX These starches are useful for all food and non-food applications of

XX starch. The present sequence is used in the course of the invention

XX SQ Sequence 1915 BP; 365 A; 610 C; 641 G; 299 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 1915;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGCGGTGCCGACCACT 289

Db 1652 GCCGCGGTGCCGACCACT 1634

RESULT 20

ACC44560/c

ID ACC44560 standard; cDNA; 2223 BP.

XX AC ACC44560;

XX DT 02-JUN-2003 (first entry)

XX DE Maize 797GL3 granule-bound starch synthase encoding cDNA SEQ ID NO:9.

XX KW Self-processing plant; plant; processing enzyme; alpha-amylase; grain;

XX KW pullulanase; alpha-glucosidase; glucose isomerase; glucoamylase;

XX KW mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;

XX KW maltodextrin; ethanol; fermentation; beverage; enzyme; gene; ss.

XX OS Zea mays.

XX Key Location/Qualifiers

XX CDS 1..2223

XX FT /\*tag= a

XX FT /partial

XX FT /product= "797GL3 granule-bound starch synthase (waxy)"

XX FT /note= "no stop codon given"

XX WO2003018766-A2.

XX PD 06-MAR-2003.

XX PF 27-AUG-2002; 2002WO-US027129.

XX PR 27-AUG-2001; 2001US-0315281P.

XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX PI Lanahan MB, Basu SS, Batie CJ, Chen W, Craig J, Kinkema M;

XX WPI; 2003-268420/26.

XX DR P-PSDB; ABP96606.

XX PT Novel polynucleotide encoding hyperthermophilic processing enzymes e.g.

XX PT alpha-amylase, useful for producing plant to produce food products having

XX PT improved taste or fermentable substrates for ethanol.

XX Claim 1; Page 92-93; 158pp; English.

XX The present invention describes polynucleotides which encode processing

XX enzymes (e.g. alpha-amylase, pullulanase, alpha-glucosidase, glucose

XX isomerase, or glucoamylase) that are optimised for expression in plants.

XX The polynucleotides encode mesophilic, thermophilic or hyperthermophilic

XX processing enzymes, which are activated under suitable conditions to act

XX upon the desired substrate. Also described are self-processing transgenic

XX plants and plant parts, e.g. grain, which express one or more of these

XX enzymes and have an altered composition that facilitates plant and grain

XX processing. Also described is a method (M) for converting starch to

XX starch-derived products in a transformed plant part (TPP), by activating

XX the starch processing enzyme contained in it. Transgenic grain is useful

XX for preparing maltodextrin. A transformed plant (TP) can be used to

XX produce food products having improved taste and to produce fermentable

XX substrates for ethanol and fermented beverages. (M) eliminates the need

XX to mill or physically disrupt the integrity of plant parts prior to

XX recovery of starch-derived products. The present sequence encodes maize

XX 797GL3 granule-bound starch synthase (waxy), which is given in the

XX exemplification of the present invention

XX SQ Sequence 2223 BP; 435 A; 729 C; 712 G; 347 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 7; Length 2223;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGCGGTGCCGACCACT 289

Db 2060 GCCGCGGTGCCGACCACT 2042

RESULT 21  
ID: AAX63355 standard; cDNA; 2267 BP.  
XX AC AAX63355;  
XX DT 16-JUL-1999 (first entry)  
XX DE Granule bound starch synthase encoding cDNA.  
XX KW Maize; corn; Zea mays; delta-9 desaturase; GBSS; target; substrate;  
KW granule bound starch synthase; hammerhead ribozyme; hairpin ribozyme;  
KW modulation; gene expression; transgenic plant; cleavage; canola plant;  
KW caffeine synthesis; coffee plant; nicotine production; tobacco;  
KW fruit ripening; flower pigmentation; lignin production; ss.  
XX OS Zea mays.  
XX PN WO9710328-A2.  
XX PD 20-MAR-1997.  
XX PF 12-JUL-1996; 96WO-US011689.  
XX PR 13-JUL-1995; 95US-0001135P.  
XX PA (RIBO-) RIBOZYME PHARM INC.  
XX FA (DOWC) DOWELANCO.  
XX PI Zwick MG, Edington BE, Mcswiggen JA, Merlo PAO, Guo L, Skokut TA;  
PI Young SA, Folkerts O, Merlo DJ;  
XX WPI; 1997-202224/18.  
XX DR Ribozyme which modulates plant gene expression - preferably modulates  
PT expression of DELTA-9 desaturase or granule bound starch synthase in  
PT maize or canola.  
XX Example 9; Page 31-33; 155pp; English.  
XX The present invention describes an enzymatic nucleic acid molecule (I)  
CC with RNA cleaving activity, which modulates the expression of a plant  
CC gene. Also described is a gene comprising a cDNA sequence encoding maize  
CC Delta-9 desaturase. (I) can be used to modulate expression of a gene,  
CC preferably Delta-9 desaturase or a granule bound starch synthase (GBSS)  
CC gene, in a plant (preferably a maize or canola plant). (I) can be used to  
CC modulate caffeine synthesis in a coffee plant, nicotine production in a  
CC tobacco plant, fruit ripening processes in an apple, tomato, pear, plum  
CC or peach plant, flower pigmentation in a rose, petunia, chrysanthemum or  
CC marigold plant or lignin production in a tobacco, aspen, poplar or pine  
CC plant  
XX SQ Sequence 2267 BP; 465 A; 699 C; 733 G; 370 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 2; Length 2267;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 271 GCCGGCGTGGCCACCACT 289  
DB 1840 GCCGGCGTGGCCACCACT 1822  
RESULT 22  
ACA25719  
ID ACA25719 standard; DNA; 3738 BP.  
XX AC ACA25719;  
XX DT 19-JUN-2003 (first entry)  
XX

DE Prokaryotic essential gene #7376.  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX Burkholderia fungorum.  
XX WO200277183-A2.  
XX PN 03-OCT-2002.  
XX PD 21-MAR-2002; 2002WO-US009107.  
XX PF 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX DR P-PSDB; ABU21849.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 13589; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 3738 BP; 690 A; 1229 C; 1290 G; 529 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 7; Length 3738;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 473 CGGACGCGCGCCGTCGA 491  
|||||

```
Db      2387 CGGACGGCGCCCGGTGGA 2405
RESULT 23
AAV29752/c
ID      AAV29752_standard; DNA; 4800 BP.
XX
XX
AC      AAV29752;
XX
XX      11-SEP-1998 (first entry)
DT
DE      Zea mays waxy gene.
XX
XX      SER; starch-encapsulating region; fusion vector; glucosyl transferase;
KW      ss.
XX
XX      Zea mays.
OS
XX      Key
FH      repeat_region
FT      283..287
FT      /tag= a
FT      /note= "direct repeat 1"
FT      repeat_region
FT      288..292
FT      /tag= b
FT      /note= "direct repeat 1"
FT      repeat_region
FT      293..297
FT      /tag= c
FT      /note= "direct repeat 1"
FT      repeat_region
FT      298..302
FT      /tag= d
FT      /note= "direct repeat 1"
FT      misc_feature
FT      372..385
FT      /tag= e
FT      /note= "GC stretch, potential regulatory factor binding"
FT      misc_feature
FT      442..468
FT      /tag= f
FT      /note= "GC stretch, potential regulatory factor binding"
FT      misc_feature
FT      768..782
FT      /tag= g
FT      /note= "GC stretch, potential regulatory factor binding"
FT      misc_feature
FT      810..822
FT      /tag= h
FT      /note= "GC stretch, potential regulatory factor binding"
FT      misc_feature
FT      821..828
FT      /tag= i
FT      /note= "target duplication site (ac7)"
FT      CAAT_signal
FT      821..828
FT      /tag= j
FT      TATA_signal
FT      867..873
FT      /tag= k
FT      misc_feature
FT      887..900
FT      /tag= l
FT      /note= "GC stretch, potential regulatory factor"
FT      exon
FT      901..1080
FT      /tag= o
FT      /number= 1
FT      misc_feature
FT      901
FT      /tag= m
FT      /note= "transcriptional start site"
FT      intron
FT      1081..1219
FT      /tag= p
FT      /number= 1
FT      exon
FT      1220..1553
FT      /tag= q
FT      /number= 2
FT      CDS
FT      1449..14343
FT      /tag= n
FT      /product= "glucosyl transferase"
FT      /note= "contains introns"
FT      intron
FT      1554..1684
FT      /tag= r
FT      /number= 2
FT      exon
FT      1685..1765
FT      /tag= s
FT      /number= 3
FT      intron
FT      1766..1859
FT      /tag= t
FT      /number= 3
FT      exon
FT      1860..1958
FT      /tag= u
FT      /number= 4
FT      intron
FT      1959..2054
FT      /tag= v
FT      /number= 4
FT      exon
FT      2055..2144
FT      /tag= w
FT      /number= 5
FT      intron
FT      2145..2225
FT      /tag= x
FT      /number= 5
FT      exon
FT      2226..2289
FT      /tag= y
FT      /number= 6
FT      intron
FT      2290..2412
FT      /tag= z
FT      /number= 6
FT      exon
FT      2413..2513
FT      /tag= aa
FT      /number= 7
FT      intron
FT      2514..2650
FT      /tag= bb
FT      /number= 7
FT      exon
FT      2651..2760
FT      /tag= cc
FT      /number= 8
FT      intron
FT      2761..2857
FT      /tag= dd
FT      /number= 8
FT      exon
FT      2858..3101
FT      /tag= ee
FT      /number= 9
FT      intron
FT      3102..3211
FT      /tag= ff
FT      /number= 9
FT      exon
FT      3212..3394
FT      /tag= gg
FT      /number= 10
FT      misc_feature
FT      3358..3489
FT      /tag= hh
FT      /note= "target duplication site (ac9)"
FT      intron
FT      3395..3489
FT      /tag= ii
FT      /number= 10
FT      exon
FT      3490..3681
FT      /tag= jj
FT      /number= 11
FT      misc_feature
FT      3570..3572
FT      /tag= kk
FT      /note= "target duplication site (spm 18)"
FT      intron
FT      3682..3792
FT      /tag= ll
FT      /number= 11
FT      exon
FT      3793..3879
FT      /tag= mm
FT      /number= 12
FT      intron
FT      3880..3976
FT      /tag= nn
FT      /number= 12
FT      exon
FT      3977..4105
FT      /tag= oo
FT      /number= 13
FT      intron
FT      4106..4226
FT      /tag= pp
FT      /number= 13
FT      exon
FT      4227..4595
FT      /tag= qq
```



PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234233P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234957P.  
PR 26-SEP-2000; 2000US-0234988P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249254P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX Disclosure; SEQ ID NO 37524; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK67694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 6029 BP; 1561 A; 1439 C; 1498 G; 1531 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 4; Length 6029;  
Best Local Similarity 100.0%; Pred.No. 29;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 274 GCGGTGCGGACCACTACC 292  
Db 5783 GCGGTGCGGACCACTACC 5801  
RESULT 26  
ID AAA63350/c  
XX AAA63350 standard; DNA; 21185 BP.  
AC AAA63350;  
XX  
DT 06-MAR-2001 (first entry)  
XX  
DE Streptomyces globisporus C-1027 gene cluster ORF 25-42.  
XX  
KW Enediyne C-1027 biosynthesis gene cluster; apoprotein; chromophore;  
KW cancer; ds.  
XX  
OS Streptomyces globisporus.  
FH Key Location/Qualifiers  
FT CDS complement(1..632)  
FT FT  
FT FT  
FT CDS /product= "type II NRPS adenylation enzyme"  
FT 1966..4044  
FT FT  
FT FT  
FT CDS /product= "transmembrane transport protein"  
FT 4188..5192  
FT FT  
FT CDS /product= "O-methyl transferase"  
FT 5249..6505  
FT FT  
FT CDS /tag= d  
FT /product= "P450 hydroxylase"  
FT complement(5539..10361)  
FT FT  
FT FT  
FT CDS /tag= g  
FT /product= "oxidoreductase"  
FT complement(6628..7734)  
FT FT  
FT CDS /tag= e  
FT /product= "oxidoreductase"  
FT 8370..9410  
FT FT  
FT CDS /tag= f  
FT /product= "ORF 31 protein"  
FT 11261..12094  
FT FT  
FT CDS /tag= h  
FT /product= "ORF 33 protein"  
FT 12251..13399  
FT FT  
FT CDS /tag= i  
FT /product= "ORF 34 protein"  
FT 14046..14900  
FT FT  
FT CDS /tag= j  
FT /product= "proline oxidase"  
FT complement(14947..15756)  
FT FT  
FT CDS /tag= k  
FT /product= "ORF 36 protein"  
FT complement(15853..16323)  
FT FT

FT FT  
FT CDS /tag= 1  
FT /product= "ORF 37 protein"  
FT complement(16460..18115)  
FT FT  
FT CDS /tag= m  
FT /product= "P450 hydroxylase"  
FT complement(18112..18642)  
FT FT  
FT CDS /tag= n  
FT /product= "ORF 39 protein"  
FT 18960..20039  
FT FT  
FT CDS /tag= o  
FT /product= "ORF 40 protein"  
FT 20065..20919  
FT FT  
FT CDS /tag= p  
FT /product= "ORF 41 protein"  
FT complement(20807..21185)  
FT FT  
FT CDS /tag= q  
FT /product= "ORF 42 protein"  
XX  
PN WO200040596-A1.  
XX  
XX 13-JUL-2000.  
XX  
XX 06-JAN-2000; 2000WO-US0000446.  
XX  
XX 06-JAN-1999; 99US-0115434P.  
PR 05-JAN-2000; 2000US-00477962.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Shen B, Liu W, Christenson SD, Standage S;  
XX  
XX WPI; 2000-465947/40.  
DR P-PSDB; AAB13588, AAB13589, AAB13590, AAB13591, AAB13592, AAB13593,  
DR AAB13594, AAB13595, AAB13596, AAB13597, AAB13598, AAB13600, AAB13601,  
DR AAB13602, AAB13603, AAB13607, AAB13606.  
XX  
PT Isolated nucleic acid comprising a nucleic acid encoding any of C-1027  
PT open reading frames (ORFs) -7 to 42, excluding ORF 9 (caga), useful for  
PT the production of enediyne C-1027 antitumor antibiotics.  
XX  
PS Claim 1; Page 130-157; 160pp; English.  
XX  
CC The present sequence is the last 21184 bases of the enediyne C-1027 gene  
CC cluster from Streptomyces globisporus. Enediyne C-1027 is an antibiotic,  
CC consisting of an apoprotein and a non-peptidic chromophore, which acts by  
CC damaging DNA. The sequences within the gene cluster, and the proteins  
CC they encode, can be used in the treatment of cancer, along with  
CC antagonists of the protein. Each of the open reading frames is  
CC specifically claimed, excluding ORF 9, which encodes Caga  
XX  
SQ Sequence 21185 BP; 2903 A; 7529 C; 7587 G; 3166 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 3; Length 21185;  
Best Local Similarity 100.0%; Pred.No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 20 CGACCGGGCGGCCAAGG 38  
Db 136 CGACCGGGCGGCCAAGG 118  
RESULT 27  
AAT93095  
ID AAT93095 standard; cDNA; 24379 BP.  
XX  
XX AAT93095;  
XX  
DT 11-MAY-1998 (first entry)  
XX  
DE Streptomyces frenolicin gene cluster.  
XX  
KW Frenolicin; antibiotic; feed additive; anticoccidial; coccidiostatic;  
KW efflux pump; butyrate starter synthase; polyketide synthase; PKS;

```

KW hemiketalase; ketoreductase; cyclase; dehydrase; ketoreductase;
KW hydroxylase; Streptomyces roseofulvus; ds.
XX Streptomyces sp.
FH Key
FT CDS      Location/Qualifiers
FT          636..2948
FT          /*tag= a
FT          /product= "80 kDa non-membrane protein"
FT          /note= "gene A (specifically claimed)"
FT          2945..3916
FT          /*tag= b
FT          /product= "membrane protein"
FT          /note= "gene B (specifically claimed)"
FT          4020..4844
FT          /*tag= c
FT          /product= "protein with 6 membrane-spanning domains"
FT          /note= "gene C (specifically claimed)"
FT          4841..6415
FT          /*tag= d
FT          /product= "ATP-binding component of ABC transporter"
FT          /note= "gene D (specifically claimed)"
FT          6533..7183
FT          /*tag= e
FT          /product= "unknown non-membrane protein"
FT          /note= "gene E (specifically claimed)"
FT          7344..8897
FT          /*tag= f
FT          /product= "putative efflux pump"
FT          /note= "gene F (specifically claimed)"
FT          9164..10012
FT          /*tag= g
FT          /product= "transcription activator"
FT          /note= "gene G (specifically claimed)"
FT          complement(10105..10621)
FT          /*tag= h
FT          /product= "translationally coupled to gene I"
FT          /note= "gene H (specifically claimed)"
FT          complement(10618..11628)
FT          /*tag= i
FT          /product= "homologue of fabH"
FT          /note= "gene I (specifically claimed)"
FT          11809..12066
FT          /*tag= j
FT          /product= "acyl carrier protein"
FT          /note= "gene J (specifically claimed)"
FT          complement(12154..13209)
FT          /*tag= k
FT          /product= "acyltransferase"
FT          /note= "gene K (specifically claimed)"
FT          13409..14686
FT          /*tag= l
FT          /product= "PKS ketoacylsynthase subunit"
FT          /note= "gene L (specifically claimed)"
FT          14767..16047
FT          /*tag= m
FT          /product= "PKS chain length factor"
FT          /note= "gene M (specifically claimed)"
FT          16120..16371
FT          /*tag= n
FT          /product= "acyl carrier protein used by the PKS"
FT          /note= "gene N (specifically claimed)"
FT          complement(16453..16935)
FT          /*tag= o
FT          /product= "putative hemiketal dehydrase"
FT          /note= "gene O (specifically claimed)"
FT          17088..17903
FT          /*tag= p
FT          /product= "ketoreductase related to actIII"
FT          /note= "gene P (specifically claimed)"
FT          17903..18898
FT          /*tag= q
FT          /product= "cyclase/dehydrase related to act VII"
FT          /note= "gene Q (specifically claimed)"
FT          18895..19839
FT          /*tag= r
FT          /product= "cyclase/dehydrase related to actIV"
FT          /note= "gene R (specifically claimed)"
FT          complement(19990..20907)
FT          /*tag= s
FT          /product= "oxidoreductase"
FT          /note= "gene S (specifically claimed)"
FT          complement(20904..22094)
FT          /*tag= t
FT          /product= "quinone-forming hydroxylase"
FT          /note= "gene T (specifically claimed)"
FT          complement(22505..22179)
FT          /*tag= u
FT          /product= "not specified"
FT          /note= "gene U"
XX
FN EP806480-A2.
XX
XX 12-NOV-1997.
XX
XX 02-MAY-1997; 97EP-00107329.
XX
XX 07-MAY-1996; 96US-0016753P.
XX
XX (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
XX Reeves CD, Soliday CL;
XX
XX WPI; 1997-538619/50.
XX
XX P-PSDB; AAW34199, AAW34200, AAW34201, AAW34202, AAW34203, AAW34204,
XX AAW34205, AAW34206, AAW34207, AAW34208, AAW34209, AAW34210, AAW34211,
XX AAW34212, AAW34213, AAW34214, AAW34215, AAW34216, AAW34217, AAW34218,
XX AAW34219.
XX
XX Streptomyces frenolicin gene cluster - useful for producing recombinant
XX frenolicin antibiotics.
XX
XX Claim 1; Page 40-60; 66pp; English.
XX
XX This DNA sequence comprises the Streptomyces frenolicin gene cluster
XX containing specifically claimed coding sequences (genes A-U) that
XX respectively encode 21 proteins (see AAW34199-219) involved in frenolicin
XX synthesis. The genes can be divided into 5 subclusters: (1) genes A, B,
XX C, D and F encode an efflux pump; (2) genes H, I, J and K encode butyrate
XX starter synthases; (3) genes L, M and N encode polyketide synthases (PKS)
XX ; (4) genes O, P, Q and R encode a hemiketalase, a ketoreductase and
XX cyclases/dehydrases; and (5) genes S and T encode a keto/enoyl reductase
XX and a hydrolase. Also claimed are vectors, host cells (especially a
XX Streptomyces sp., particularly Streptomyces roseofulvus), and the encoded
XX proteins. Cells transformed using the above sequence can be cultured to
XX produce frenolicins or frenolicin precursors. The precursors can be
XX converted to frenolicins by chemical or other methods. The frenolicins
XX can be oxidised to frenolicin B, an antibiotic used as an anticoccidial
XX agent. The frenolicins can be used as animal feed additives
XX
XX SQ Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 U; 0 Other;
XX
XX Query Match 3.8%; Score 19; DB 2; Length 24379;
XX Best Local Similarity 100.0%; Pred. No. 24;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 477 CGGCGCCGCGTCGAGGCG 495
XX |||||
XX Db 17279 CGGCGCCGCGTCGAGGCG 17297
XX
XX RESULT 28
XX AAV25925
XX ID AAV25925 standard; cDNA; 24379 BP.
XX
XX AC AAV25925;
```



```

XX 15-JUL-1998 (first entry)
XX Streptomyces roseofulvus frenolicin gene cluster.
XX Streptomyces roseofulvus; frenolicin gene cluster; frenolicin B;
XX antibiotic; ss.
XX Streptomyces roseofulvus.
XX Location/Qualifiers
XX 636..2948
XX CDS
XX /tag= a
XX /note= "encodes protein given in AAW55800"
XX 2945..3916
XX /tag= b
XX /transl_except= (pos:2945..2947,aa:Met)
XX /note= "encodes protein given in AAW55801"
XX 4020..4844
XX /tag= c
XX /transl_except= (pos:4020..4022,aa:Met)
XX /note= "encodes protein given in AAW55802"
XX 4841..6415
XX /tag= d
XX /note= "encodes protein given in AAW55803"
XX 6533..7183
XX /tag= e
XX /transl_except= (pos:6533..6535,aa:Met)
XX /note= "encodes protein given in AAW55804"
XX 7344..8897
XX /tag= f
XX /note= "encodes protein given in AAW55805"
XX 9164..10012
XX /tag= g
XX /transl_except= (pos:9164..9166,aa:Met)
XX /note= "encodes protein given in AAW55806"
XX complement(10105..10621)
XX /tag= h
XX /transl_except= (pos:10621..10619,aa:Met)
XX /note= "encodes protein given in AAW55807"
XX complement(10618..11628)
XX /tag= i
XX /transl_except= (pos:11628..11626,aa:Met)
XX /note= "encodes protein given in AAW55808"
XX 11809..12066
XX /tag= j
XX /note= "encodes protein given in AAW55809"
XX complement(12154..13209)
XX /tag= k
XX /transl_except= (pos:13209..13207,aa:Met)
XX /note= "encodes protein given in AAW55810"
XX 13409..14686
XX /tag= l
XX /transl_except= (pos:13409..13411,aa:Met)
XX /note= "encodes protein given in AAW55811"
XX 14767..16047
XX /tag= m
XX /note= "encodes protein given in AAW55812"
XX 16120..16371
XX /tag= n
XX /note= "encodes protein given in AAW55813"
XX complement(16453..16935)
XX /tag= o
XX /transl_except= (pos:16935..16933,aa:Met)
XX /note= "encodes protein given in AAW55814"
XX 17088..18903
XX /tag= p
XX /note= "encodes protein given in AAW55815"
XX 17903..18898
XX /tag= q
XX /note= "encodes protein given in AAW55816"
XX 18895..19839

```

```

FT /tag= i
FT /note= "encodes protein given in AAW55817"
FT complement(19990..20907)
FT /tag= s
FT /transl_except= (pos:20907..20905,aa:Met)
FT /note= "encodes protein given in AAW55818"
FT complement(20904..22094)
FT /tag= t
FT /transl_except= (pos:22040..22038,aa:Arg)
FT /transl_except= (pos:22037..22035,aa:Arg)
FT /transl_except= (pos:22034..22032,aa:Thr)
FT /transl_except= (pos:22031..22029,aa:Leu)
FT /note= "encodes protein given in AAW55819"
XX
XX JP10094395-A.
XX 14-APR-1998.
XX 07-MAY-1997; 97JP-00116652.
XX 07-MAY-1996; 96US-0016753P.
XX 04-APR-1997; 97US-0042935P.
XX (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX WPI; 1998-279231/25.
XX P-PSDB; AAW55800, AAW55801, AAW55802, AAW55803, AAW55804, AAW55805,
XX AAW55806, AAW55807, AAW55808, AAW55809, AAW55810, AAW55811, AAW55812,
XX AAW55813, AAW55814, AAW55815, AAW55816, AAW55817, AAW55818, AAW55819.
XX Frenolicin gene cluster - useful as an antibiotic.
XX Claim 2; Page 36-44; 50pp; Japanese.
XX The present sequence represents a frenolicin gene cluster from
XX Streptomyces roseofulvus. The present invention describes: (1) a vector
XX containing the frenolicin gene cluster DNA sequence operably connected to
XX an expression control sequence; (2) a host cell transformed by the above
XX vector; (3) a protein coded by the above DNA sequence; (4) a method for
XX the preparation of frenolicin or a biosynthetic intermediate for it in
XX which the above cell is cultured and frenolicin or its biosynthetic
XX intermediate is isolated from the culture or the cell; (5) a method for
XX the preparation of frenolicin B by oxidising frenolicin; and (6) a method
XX for the preparation of a feed composition by mixing frenolicin with other
XX components. Frenolicin B is useful as an antibiotic
XX Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 U; 0 Other;
XX Query Match 3.8%; Score 19; DB 2; Length 24379;
XX Best Local Similarity 100.0%; Pred.No.24;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 477 CGGCGCCCGCGTCGAGGCG 495
XX |
XX 17279 CGGCGCCCGCGTCGAGGCG 17297
XX
XX RESULT 29
XX AAA63348/c
XX ID AAA63348 standard; DNA; 63164 BP.
XX AC
XX AAA63348;
XX AC
XX 06-MAR-2001 (first entry)
XX Streptomyces globisporus C-1027 gene cluster.
XX Eneidiyne C-1027 biosynthesis gene cluster; apoprotein; chromophore;
XX cancer; ds.
XX Streptomyces globisporus.
XX Key Location/Qualifiers

```

FT	CDS	complement(8..658)
FT	FT	/tag= a
FT	FT	/product= "ORF -7 protein"
FT	CDS	complement(930..1478)
FT	FT	/tag= b
FT	FT	/product= "ORF -6 protein"
FT	CDS	complement(1649..2713)
FT	FT	/tag= c
FT	FT	/product= "ORF -5 protein"
FT	CDS	complement(2850..3237)
FT	FT	/tag= d
FT	FT	/product= "ORF -4 protein"
FT	CDS	complement(3442..4971)
FT	FT	/tag= e
FT	FT	/product= "ORF -3 protein"
FT	CDS	5982..7479
FT	FT	/tag= f
FT	FT	/product= "glycerol phosphate transporter"
FT	CDS	complement(7573..9900)
FT	FT	/tag= g
FT	FT	/product= "ABC transport/UvrA-like protein"
FT	CDS	complement(9982..11349)
FT	FT	/tag= h
FT	FT	/product= "Na+/H+ transporter"
FT	CDS	complement(11351..12835)
FT	FT	/tag= i
FT	FT	/product= "hydroxylase/halogenase"
FT	CDS	13012..14079
FT	FT	/tag= j
FT	FT	/product= "dNDP-glucose synthase"
FT	CDS	complement(14212..14643)
FT	FT	/tag= k
FT	FT	/product= "CagA"
FT	CDS	complement(14690..15922)
FT	FT	/tag= l
FT	FT	/product= "aminotransferase"
FT	CDS	complement(15919..16653)
FT	FT	/tag= m
FT	FT	/product= "N-methyl transferase"
FT	CDS	complement(16653..17924)
FT	FT	/tag= n
FT	FT	/product= "C-methyl transferase"
FT	CDS	complement(18031..19191)
FT	FT	/tag= o
FT	FT	/product= "epoxide hydrazse"
FT	CDS	complement(19267..19929)
FT	FT	/tag= p
FT	FT	/product= "anthranilate synthase II"
FT	CDS	complement(19926..21407)
FT	FT	/tag= q
FT	FT	/product= "anthranilate synthase I"
FT	CDS	complement(21424..22878)
FT	FT	/tag= r
FT	FT	/product= "coenzyme F390 synthetase"
FT	CDS	complement(22875..23546)
FT	FT	/tag= s
FT	FT	/product= "iron-sulphur flavoprotein"
FT	CDS	complement(23566..24702)
FT	FT	/tag= t
FT	FT	/product= "O-acyl transferase"
FT	CDS	complement(24986..25564)
FT	FT	/tag= u
FT	FT	/product= "epimerase"
FT	CDS	25815..27170
FT	FT	/tag= v
FT	FT	/product= "monooxygenase"
FT	CDS	27214..28593
FT	FT	/tag= w
FT	FT	/product= "glycosyl transferase"
FT	CDS	28590..29588
FT	FT	/tag= x
FT	FT	/product= "dNDP glucose dehydratase"
FT	CDS	29632..31197
FT	FT	/tag= y
FT	FT	/product= "SgcB transmembrane efflux protein"
FT	CDS	31280..32590
FT	FT	/tag= z
FT	FT	/product= "coenzyme F390 synthetase"
FT	CDS	32809..34392
FT	FT	/tag= aa
FT	FT	/product= "hydroxylase"
FT	CDS	complement(34458..35294)
FT	FT	/tag= ab
FT	FT	/product= "ORF 22 protein"
FT	FT	/transl_except= {pos:35223..35225,aa:Ile}
FT	FT	/transl_except= {pos:35226..35228,aa:Ala}
FT	CDS	complement(35518..35938)
FT	FT	/tag= ac
FT	FT	/product= "ORF 18 protein"
FT	CDS	36165..37490
FT	FT	/tag= ad
FT	FT	/product= "coenzyme F390 synthetase"
FT	CDS	37559..38938
FT	FT	/tag= ae
FT	FT	/product= "type II NRPS condensation enzyme"
FT	CDS	38983..39264
FT	FT	/tag= af
FT	FT	/product= "type II peptidyl carrier protein"
FT	CDS	complement(39367..40986)
FT	FT	/tag= ag
FT	FT	/product= "aminomutase"
FT	CDS	complement(41052..42611)
FT	FT	/tag= ah
FT	FT	/product= "type II NRPS adenylation enzyme"
FT	CDS	43945..46024
FT	FT	/tag= ai
FT	FT	/product= "transmembrane transport protein"
FT	CDS	46167..47171
FT	FT	/tag= aj
FT	FT	/product= "O-methyl transferase"
FT	CDS	47228..48484
FT	FT	/tag= ak
FT	FT	/product= "P450 hydroxylase"
FT	CDS	complement(48607..49714)
FT	FT	/tag= al
FT	FT	/product= "oxidoreductase"
FT	CDS	50350..51390
FT	FT	/tag= am
FT	FT	/product= "ORF 31 protein"
FT	CDS	complement(51421..52341)
FT	FT	/tag= an
FT	FT	/product= "oxidoreductase"
FT	CDS	53241..54074
FT	FT	/tag= ao
FT	FT	/product= "ORF 33 protein"
FT	CDS	54231..55379
FT	FT	/tag= ap
FT	FT	/product= "ORF 34 protein"
FT	CDS	56026..56880
FT	FT	/tag= aq
FT	FT	/product= "proline oxidase"
FT	CDS	complement(56927..57736)
FT	FT	/tag= ar
FT	FT	/product= "ORF 36 protein"
FT	CDS	complement(57833..58304)

```
FT CDS 62045..62899
FT     /*tag= aw
FT     /product= "ORF 41 protein"
FT CDS complement(62787..63164)
FT     /*tag= ax
FT     /product= "ORF 42 protein"
XX
XX WO200040596-A1.
XX
XX 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-US000446.
XX
XX 06-JAN-1999; 99US-0115434P.
XX
XX 05-JAN-2000; 2000US-00477962.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Shen B, Liu W, Christenson SD, Standage S;
XX
XX WPI; 2000-465947/40.
XX
XX P-PSDB; AAB13554, AAB13555, AAB13556, AAB13557, AAB13558, AAB13559,
XX AAB13560, AAB13561, AAB13562, AAB13563, AAB13564, AAB13565, AAB13566,
XX AAB13567, AAB13568, AAB13569, AAB13570, AAB13571, AAB13572, AAB13573,
XX AAB13574, AAB13575, AAB13576, AAB13577, AAB13578, AAB13579, AAB13580,
XX AAB13581, AAB13582, AAB13583, AAB13584, AAB13585, AAB13586, AAB13587,
XX AAB13588, AAB13589, AAB13590, AAB13591, AAB13592, AAB13593, AAB13594,
XX AAB13595, AAB13596, AAB13597, AAB13598, AAB13599, AAB13600, AAB13601, AAB13602,
XX AAB13603, AAB13606, AAB13607.
XX
XX Isolated nucleic acid comprising a nucleic acid encoding any of C-1027
XX open reading frames (ORFs) -7 to 42, excluding ORF 9 (caga), useful for
XX the production of enediyme C-1027 antitumor antibiotics.
XX
XX Claim 1; Page 78-157; 160pp; English.
XX
XX The present invention is the enediyme C-1027 gene cluster from
XX Streptomyces globisporus. Enediyme C-1027 is an antibiotic, consisting of
XX an apoprotein and a non-peptidic chromophore, which acts by damaging DNA.
XX The sequences within the gene cluster, and the proteins they encode, can
XX
XX Query Match 3.8%; Score 19; DB 3; Length 63164;
XX Best Local Similarity 100.0%; Pred. No. 22;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 20 CGACCGGGCGGCCCAAGG 38
XX |||||
XX Db 42115 CGACCGGGCGGCCCAAGG 42097
XX
XX RESULT 30
XX ABQ81846/c
XX ID ABQ81846 standard; DNA; 349980 BP.
XX
XX AC ABQ81846;
XX
XX 19-NOV-2002 (first entry)
XX
XX Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1102.
XX
XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
XX antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
XX identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
XX rotavirus; food composition; pharmaceutical composition; gene; db.
XX
XX Bifidobacterium longum.
XX OS Synthetic.
XX
XX EP1227152-A1.
XX
XX 31-JUL-2002.
XX
XX 30-JAN-2001; 2001EP-00102050.
XX
XX 30-JAN-2001; 2001EP-00102050.
XX
XX (NEST ) SOC PROD NESTLE SA.
XX
XX WPI; 2002-668397/72.
XX
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as
XX a probe or primer for detecting and/or identifying Bifidobacterium longum
XX in a biological sample.
XX
XX Disclosure; SEQ ID NO 1102; 80pp; English.
XX
XX The present invention describes a polynucleotide (I) comprising a
XX sequence of a Bifidobacterium genome selected from the nucleotide
XX sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
XX least 90% identity or which hybridises with the sequences given in
XX ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a
XX fusion protein, comprising a sequence selected from 1097 sequences given
XX in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a
XX heterologous polypeptide. (I) has antidiarrheic and antibacterial
XX activities, and can be used as an inhibitor of Salmonella. (I) (which is
XX a probe) is useful for the detection and/or identification of
XX Bifidobacterium longum in a biological sample. A carrier containing the
XX lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM 1-2618) can be
XX used for preventing and/or treating diarrhoea brought about by pathogenic
XX bacteria and/or rotavirus. The carrier is a food composition selected
XX from milk, yogurt, curd, cheese, fermented milks, milk based fermented
XX products, ice-creams, fermented cereal based products, milk based
XX powders, infant formula, pet food or a pharmaceutical composition
XX selected from tablets, liquid bacterial suspensions, dried oral
XX supplement, wet oral supplement, dry tube feeding or wet tube feeding.
XX (I) is useful in DNA arrays or chips to carry out analysis of the
XX expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
XX Bifidobacterium related nucleotide sequences given in the Sequence
XX listing from the present invention but not mentioned further within the
XX specification. N.B. The sequence data for this patent is not represented
XX in the printed specification but is based on sequence information
XX supplied by the European Patent Office
XX
XX Sequence 349980 BP; 69200 A; 103414 C; 105147 G; 72219 T; 0 U; 0 Other;
XX
XX Query Match 3.8%; Score 19; DB 6; Length 349980;
XX Best Local Similarity 100.0%; Pred. No. 18;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 219 CTCGCGCGACGACGTGATG 237
XX |||||
XX Db 247972 CTCGCGCGACGACGTGATG 247954
XX
XX RESULT 31
XX AAI29319
XX ID AAI29319 standard; cDNA; 175 BP.
XX
XX AC AAI29319;
XX
XX 12-OCT-2001 (first entry)
XX
XX Colon tumour related determined cDNA sequence for clone R0095:F10.
XX
XX Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
XX gene therapy; vaccine; colonic cancer; ss.
XX
XX Homo sapiens.
XX OS
XX WO200149716-A2.
XX
XX 12-JUL-2001.
XX
XX 29-DEC-2000; 2000WO-US035596.
XX
XX 30-DEC-1999; 99US-00476296.
XX
```

```

PR 10-JAN-2000; 2000US-00480321.
PR 15-FEB-2000; 2000US-00504629.
PR 06-MAR-2000; 2000US-00519444.
PR 19-MAY-2000; 2000US-00575251.
PR 29-JUN-2000; 2000US-00609448.
PR 28-AUG-2000; 2000US-00649811.
XX
PA (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
PI Fanger GR, Vedvick TS, Carter D;
XX
DR WPI; 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer.
XX
XX Claim 2; Page 376; 472pp; English.
XX
XX The present invention describes colon tumor associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumor associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be used
CC to treat disorders associated with decreased expression by rectifying
CC mutations or deletions in a patient's genome that affect the activity of
CC TCAPs by expressing inactive proteins or to supplement the patients own
CC production of them. Additionally, (II) may be used to produce the TCAP
CC proteins, by inserting the nucleic acids into a host cell culturing the
CC cell to express the protein. (II) and its complementary sequences may
CC also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
CC and hybridisation assays to detect and quantitate the presence of similar
CC nucleic acids in samples, and therefore which patients may be in need of
CC restorative therapy. (I) may also be used as antigens in the production
CC of antibodies against TCAPs and in assays to identify modulators of TCAP
CC expression and activity. Anti-(I) antibodies and antagonists may also be
CC used to down regulate TCAP expression and activity. The anti-(I)
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of TCAPs in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA)). AA128460 to AA129512 and AA24494 to AA24523 represent
CC nucleotide and amino acid sequences given in the exemplification of the
CC present invention
XX
XX Sequence 175 BP; 30 A; 70 C; 56 G; 19 T; 0 U; 0 Other;
SQ
Query Match 3.6%; Score 18; DB 4; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 297 CCTGGTGGCGGCCCG 314
DB 36 CCTGGTGGCGGCCCG 53
RESULT 32
ABZ33505
ID ABZ333505 standard; cDNA; 175 BP.
XX
XX ABZ333505;
XX
XX 30-JAN-2003 (first entry)
XX
XX Human colon tumour cDNA for clone R0095.F10 SEQ ID NO:873.
XX
XX Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;
XX tumour; immune response; immunostimulant; cytostatic; vaccine; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200283070-A2.
XX
XX 24-OCT-2002.
XX
PR 09-APR-2002; 2002WO-US011475.
XX
XX 10-APR-2001; 2001US-00833263.
PR 03-AUG-2001; 2001US-00922217.
PR 19-DEC-2001; 2001US-00025380.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI Wang T, Jiang Y, Smith CL, King GE, Wang A, Clapper JD, Skeiky YAW;
PI Fanger GR, Vedvick TS, Carter D;
XX
XX WPI; 2003-067548/06.
XX
XX New polynucleotide, useful for the preparation of a composition for
PT stimulating an immune response against, or treating, cancer.
XX
XX Disclosure; Page 378; 537pp; English.
XX
XX The present invention describes compounds (I) for the immunotherapy and
CC diagnosis of colon cancer. Also described: (1) a method for detecting the
CC presence of cancer in a patient; (2) a method for stimulating and/or
CC expanding T cells specific for a tumour protein; (3) an isolated T cell
CC population comprising T cells prepared by the method of (2); (4) a method
CC for stimulating an immune response in a patient; (5) a method for
CC treating cancer in a patient; and (6) a method for inhibiting the
CC development of cancer in a patient. (I) have immunostimulant and
CC cytostatic activities and can be used in vaccines. ABZ32646 to ABZ33725
CC and ABP55343 to ABP55391 represent human colon cancer/tumour related
CC sequences used in the exemplification of the present invention
XX
XX Sequence 175 BP; 30 A; 70 C; 56 G; 19 T; 0 U; 0 Other;
SQ
Query Match 3.6%; Score 18; DB 7; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 297 CCTGGTGGCGGCCCG 314
DB 36 CCTGGTGGCGGCCCG 53
RESULT 33
ABN18020
ID ABN18020 standard; cDNA; 313 BP.
XX
XX ABN18020;
XX
XX 24-JUN-2002 (first entry)
XX
XX Human ORFX polynucleotide sequence SEQ ID NO:4517.
XX
XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX hypertension; hypothyroidism; cholesterol ester storage disease;
XX immune deficiency; immune disorder; infectious disease;
XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX myasthenia gravis; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200192523-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US010836.
XX
XX 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX

```

PA (CURA-) CURAGEN CORP.  
XX Shimkets RA, Leach MD;  
XX WPI; 2002-106308/14.  
DR P-PSDB; ABP02268.  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders.  
XX  
XX Disclosure; SEQ ID NO 4517; 1037pp; English.  
XX  
XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 313 BP; 61 A; 107 C; 87 G; 58 T; 0 U; 0 Other;  
  
Query Match 3.6%; Score 18; DB 6; Length 313;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 211 GACGCGAGCTCGCGGAC 228  
Db |||||||  
223 GACGCGAGCTCGCGGAC 240  
  
RESULT 34  
ACH19833  
ID ACH19833 standard; cDNA; 418 BP.  
XX  
XX ACH19833;  
XX  
XX  
DT 13-OCT-2003 (first entry)  
XX  
XX Human adult lung cDNA #836.  
XX  
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
KW genome mapping; biodiversity; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX US2003073623-A1.  
XX  
XX 17-APR-2003.  
XX  
XX 30-JUL-2001; 2001US-00918995.  
XX  
XX 30-JUL-2001; 2001US-00918995.  
XX  
XX (DRMA/) DRMANAC R T.  
  
(LABA/) LABAT I.  
PA (STAC/) STACHE-CRAIN B.  
PA (DICK/) DICKSON M C.  
PA (JONE/) JONES L W.  
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX WPI; 2003-615964/58.  
XX  
XX New polynucleotide sequences obtained from various cDNA libraries, useful  
XX as hybridization probes, as oligomers for PCR, for chromosome and gene  
XX mapping, in the recombinant production of protein, or in generating  
XX antisense DNA or RNA.  
XX  
XX Claim 1; SEQ ID NO 7045; 44pp; English.  
XX  
XX The invention relates to an isolated polynucleotide comprising any one of  
XX 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
XX determined by the technique of SBH (sequencing by hybridisation). Also  
XX included is a purified polypeptide comprising a sequence corresponding to  
XX a reading frame of the novel polynucleotide. The nucleic acid sequences  
XX are useful in diagnostics as expressed sequence tags (EST) for  
XX identifying expressed genes or for physical mapping of the human genome,  
XX in forensics, in assessing biodiversity, or in identifying mutations  
XX responsible for genetic disorders and other traits. The nucleotide  
XX sequences are also useful as hybridisation probes, as oligomers for PCR,  
XX for chromosome and gene mapping, in the recombinant production of  
XX protein, or in generating antisense DNA or RNA. The purified polypeptide  
XX is useful for generating antibodies specific for it. The present sequence  
XX is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
XX for this patent did not form part of the printed specification, but was  
XX obtained in electronic format directly from USPTO at  
XX seqdata.uspto.gov/sequence.html?DocID=20030073623  
XX  
SQ Sequence 418 BP; 89 A; 128 C; 133 G; 68 T; 0 U; 0 Other;  
  
Query Match 3.6%; Score 18; DB 8; Length 418;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 350 GCCTGCGCGGGGCGCGG 367  
Db |||||||  
309 GCCTGCGCGGGGCGCGG 326  
  
RESULT 35  
AAH83034  
ID AAH83034 standard; cDNA; 457 BP.  
XX  
XX AAH83034;  
XX  
XX  
DT 25-SEP-2001 (first entry)  
XX  
XX Human ovarian tumour associated polynucleotide sequence SEQ ID NO:658.  
XX Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;  
KW immunogenic; vaccine; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200151513-A2.  
XX  
XX 19-JUL-2001.  
XX  
XX 16-JAN-2001; 2001WO-US001575.  
XX  
XX 14-JAN-2000; 2000US-0176722P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Algate PA;  
XX  
XX WPI; 2001-425866/45.  
XX  
XX DR

XX		Noel ovarian tumor proteins, and nucleic acids encoding them, used to treat and diagnose cancers, particularly ovarian cancer.
PT		
FT		
PS		
CC		Claim 5; Page 185; 338pp; English.
XX		AAMH92377 to AAMH93878 represent human ovarian tumour-associated polynucleotide sequences which encode ovarian tumour proteins. The ovarian tumour protein and polynucleotide sequences have cytostatic activity, and can be used in gene therapy and vaccine production. The CC
CC		ovarian tumour proteins and polynucleotides can be used to inhibit the development of cancer, particularly ovarian cancer. They can also be used CC
CC		to diagnose the onset and progression of cancer
XX		
SQ		Sequence 457 BP; 127 A; 83 C; 87 G; 157 T; 0 U; 3 Other;
	Query Match	3.6%; Score 18; DB 5; Length 457;
	Best Local Similarity	100.0%; Pred. No. 1.2e+02;
	Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
QY	23	C CGGGCGGCGGCCAAGGGCG 40       
Dd	367	C CGGGCGGCGGCCAAGGGCG 384
RESULT 36		
AAV29373	ID	AAV29373 standard; cDNA; 498 BP.
XX AC	AAV29373;	
XX DT	DT	(first entry)
DE DE		Human small nuclear ribo-nucleoprotein Sm protein (HSMPA) encoding cDNA.
KW KW		Small nuclear ribo-nucleoprotein; snRNP; Sm protein; HSMP; human; diagnosis; prevention; treatment; systemic lupus erythematosus; ss.  
OS OS	Homo sapiens.	
PH PH	Key Location/Qualifiers	
FT FT	CDS 28..339	
FT FT	/tag= a	
XX XX	/product= "HSMPA protein"	
PN PN	WO9813380-A1.	
PD PD	02-APR-1998.	
XX XX	26-SEP-1997; 97WO-US017362.	
PF PF	27-SEP-1996; 96US-00722349.	
PR PR	(INCY-) INCYTE PHARM INC.	
PA PA	Hillman JL, Bandman O, Zweiger GB;	
PI PI	WI: 1998-230633/20.	
XX XX	P-PADB; AAWS6788.	
DR DR		New isolated human small nuclear ribonucleoprotein Sm proteins - is used PT
XX XX		to develop products for diagnosis, study, prevention and treatment of PT
XX XX		disease, particularly systemic lupus erythematosus.
PT FT		Claim 3; Fig 1A-B; 59pp; English.
PS PS		This cDNA encodes a human small nuclear ribonucleoprotein (snRNP) Sm
XX XX		protein (HSMP). A recombinant host cell containing a vector comprising
CC CC		polynucleotide sequences encoding HSMPA or HSMPPB can be used to produce
CC CC		these proteins. The HSMPA and HSMPPB and its products can be used in the
CC CC		diagnosis, prevention and treatment of systemic lupus erythematosus. In
CC CC		particular agents which modulate the activity of HSNP would be useful
CC CC		therapeutically for decreasing the harmful effects of the anti-Sm immune

CC response in systemic lupus erythematosus  
XX  
SQ Sequence 498 BP; 112 A; 143 C; 151 G; 92 T; 0 U; 0 Other;  
  
Query Match 3.6%; Score 18; DB 2; Length 498;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 350 GCCTGGCGGCGGCGGCG 367  
|||||  
DB 339 GCCTGGCGGCGGCGGCG 356  
  
RESULT 37  
AAS75981  
ID AAS75981 standard; cDNA; 537 BP.  
AC AAS75981;  
XX  
DT  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #11785.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200175067-A2.  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPT; 2001-639362/73.  
DR P-PSDE; ABG11794.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity.  
XX  
XX Claim 1; SEQ ID NO 11785; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II).  
sequences (I) is useful as hybridisation probes, polymerase chain  
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
and in recombinant production of (II). The polynucleotides are also used  
in diagnostics as expressed sequence tags for identifying expressed  
genes. (I) is useful in gene therapy techniques to restore normal  
activity of (II) or to treat disease states involving (II). (II) is  
useful for generating antibodies against it, detecting or quantitating a  
polypeptide in tissue, as molecular weight markers and as a food  
supplement. (II) and its binding partners are useful in medical imaging  
of sites expressing (II). (I) and (II) are useful for treating disorders  
involving aberrant protein expression or biological activity. The  
polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
coding sequences of the invention. Note: The sequence data for this  
patent did not appear in the printed specification, but was obtained in  
electronic format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 537 BP; 80 A; 198 C; 189 G; 68 T; 0 U; 2 Other;  
Query Match 3.6%; Score 18; DB 5; Length 537;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 297 CCTGGTGGGCGCCCG 314  
Db 327 CCTGGTGGGCGCCCG 344  
RESULT 38  
AAL01672  
ID AAL01672 standard; cDNA; 622 BP.  
XX  
XX AAL01672;  
XX AC  
XX  
DT 21-NOV-2001 (first entry)  
XX  
XX Human reproductive system related antigen cDNA SEQ ID NO: 1673.  
DE Human reproductive system related antigen; reproductive system disorder;  
XX Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ss.  
XX Homo sapiens.  
OS  
XX WO20015320-A2.  
XX  
XX  
PD 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001339.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-01845664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225577P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 22-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.

PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 05-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX WPI; 2001-465570/50.  
DR P-PSDB; AAM95702.  
DR  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
PT used in preventing, treating or ameliorating a medical condition.  
PT  
XX Claim 1; SEQ ID NO 1673; 1297pp + Sequence Listing; English.  
PS  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a coding sequence of the  
CC invention  
XX  
XX Sequence 622 BP; 181 A; 166 C; 169 G; 104 T; 0 U; 2 Other;  
SQ  
Query Match 3.6%; Score 18; DB 4; Length 622;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 350 GCCTGGCGGGGGCGCG 367  
Db 409 GCCTGGCGGGGGCGCG 426  
RESULT 39  
ABA07591  
ID ABA07591 standard; cDNA; 622 BP.  
XX  
AC ABA07591;  
XX  
XX 11-JAN-2002 (first entry)  
DT  
XX Human ovarian and breast cancer associated polynucleotide SEQ ID NO 148.  
DE  
XX Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiact; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200155325-A2.  
FN  
XX 02-AUG-2001.  
PD

XX 17-JAN-2001; 2001WO-US001345.  
PF  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 25-SEP-2000; 2000US-0234977P.  
PR 25-SEP-2000; 2000US-0234988P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.



PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0246417P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256713P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PR (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-488786/53.  
DR P-PSDB; ABB10880.  
XX

XX New isolated ovarian and/or breast cancer related nucleic acids and polypeptides, useful for diagnosing, treating and/or preventing human diseases and disorders, particularly ovarian and/or breast cancer.  
XX Claim 1; SEQ ID NO 148; 577pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABA07454-ABA08224) and proteins (ABB10743-ABB10980) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 622 BP; 181 A; 166 C; 169 G; 104 T; 0 U; 2 Other;  
Query Match 3.6%; Score 18; DB 4; Length 622;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 350 GCCTGGCGGGCGGCGG 367  
Db 409 GCCTGGCGGGCGGCGG 426  
|||||  
RESULT 40  
AAS81267  
ID AAS81267 standard; cDNA; 633 BP.  
XX  
AC AAS81267;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #17071.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABB17080.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.  
XX  
PS Claim 1; SEQ ID NO 17071; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridization probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, CC and in recombinant production of (II). The polynucleotides are also used CC in diagnostics as expressed sequence tags for identifying expressed CC genes. (I) is useful in gene therapy techniques to restore normal CC activity of (II) or to treat disease states involving (II). (II) is CC useful for generating antibodies against it, detecting or quantitating a CC polypeptide in tissue, as molecular weight markers and as a food CC supplement. (II) and its binding partners are useful in medical imaging CC of sites expressing (II). (I) and (II) are useful for treating disorders CC involving aberrant protein expression or biological activities. The CC polypeptide and polynucleotide sequences have applications in CC diagnostics, forensics, gene mapping, identification of mutations CC and to produce other types of data and products dependent on DNA and CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic CC coding sequences of the invention. Note: The sequence data for this CC patent did not appear in the printed specification, but was obtained in CC electronic format directly from WIPO at CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 633 BP; 151 A; 189 C; 181 G; 112 T; 0 U; 0 Other;

Query Match 3.6%; Score 18; DB 5; Length 633;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGGGGGGGCGG 367  
|||||  
DB 454 GCCTGGCGGGGGGGCGG 471

RESULT 41  
ACC83519  
ID ACC83519 standard; cDNA; 819 BP.  
XX  
AC ACC83519;  
XX  
DT 08-SEP-2003 (first entry)  
XX  
DE Human exosome component hMTR3 coding sequence.  
XX  
KW Human; exosome; hMTR3; mRNA turnover; antiinflammatory; cytostatic; gene;  
KW ss.  
XX Homo sapiens.  
XX  
FN WO2003044166-A2.  
XX  
PD 30-MAY-2003.  
XX  
PF 14-NOV-2002; 2002WO-US036665.  
XX  
PR 15-NOV-2001; 2001US-0334712P.  
XX (REGC ) UNIV CALIFORNIA.  
PA  
PI Chen C, Karin M;  
XX  
XX WPI; 2003-457602/43.  
DR P-PSDB; ABR42950.  
XX  
XX New purified mammalian exosome, useful for treating immune diseases e.g.  
PT diabetes, cirrhosis, scleroderma, lupus, arthritis, or multiple  
PT sclerosis, inflammatory diseases, cancer, or hormone deficiency disease  
PT e.g. osteoporosis.  
XX  
PS Disclosure; Fig 3D; 127pp; English.  
XX  
XX The present sequence is the coding sequence of the human exosome  
CC component, hMTR3. Peptide fragments (see ABR42935-38) of hMTR3 were

CC identified in a protein isolated from the human exosome. The human  
CC exosome promotes degradation of AU-rich element (ARE)-containing RNAs.  
CC ARE recognition by the exosome requires ARE-binding proteins that  
CC interact with the exosome and recruit it to unstable RNAs, thereby  
CC promoting their rapid degradation. A claimed method for altering the  
CC level of an ARE-containing RNA involves contacting the RNA with an ARE-  
CC binding protein, a mammalian (human) exosome and an agent that alters the  
CC level of binding of the exosome to the ARE-binding protein. Levels of  
CC mRNA are decreased or increased by impacting the level of degradation of  
CC the mRNA. Reducing the level of a cytokine, proto-oncogene or growth  
CC factor RNA can be used to treat an immune disease, inflammatory disease  
CC or cancer. Increasing the level of a target RNA can be performed in a  
CC subject capable of developing a hormone deficiency

XX SQ Sequence 819 BP; 87 A; 311 C; 330 G; 91 T; 0 U; 0 Other;

Query Match 3.6%; Score 18; DB 8; Length 819;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 CCTGGTGGCGGGCGCGCG 314  
|||||  
DB 774 CCTGGTGGCGGGCGCGCG 791

RESULT 42  
ACA44265/c  
ID ACA44265 standard; DNA; 909 BP.  
XX  
AC ACA44265;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene #25922.  
XX  
KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
OS Pseudomonas putida.  
XX  
FN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU40395.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 32135; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 909 BP; 188 A; 275 C; 264 G; 182 T; 0 U; 0 Other;  
Query Match 3.6%; Score 18; DB 7; Length 909;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 246 CGAGCGGAGAGCTCGAC 263  
D5 294 CGAGCGGAGAGCTCGAC 277  
|||||

RESULT 43  
AAH14082  
ID AAH14082 standard; cDNA; 1699 BP.  
XX AC AAH14082;  
XX 26-JUN-2001 (first entry)  
XX Human cDNA sequence SEQ ID NO:11239.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-00116126.  
XX 29-JUL-1999; 99JP-00248036.  
XX 27-AUG-1999; 99JP-00300253.  
XX 11-JAN-2000; 2000JP-00118776.  
XX 02-MAY-2000; 2000JP-001183767.  
XX 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
XX length cDNAs defined in the specification, and for the detection and/or  
XX diagnosis of the abnormality of the proteins encoded by the full-length  
XX cDNAs.

XX Claim 8; SEQ ID NO 11239; 2537pp + Sequence Listing; English.  
XX The present invention describes primer sets for synthesizing 5602 full-  
XX length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
XX present invention

XX SQ Sequence 1699 BP; 338 A; 556 C; 440 G; 365 T; 0 U; 0 Other;  
Query Match 3.6%; Score 18; DB 4; Length 1699;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 284 CCACCTACACACCTCGG 301  
D5 955 CCACCTACACACCTCGG 972  
|||||

RESULT 44  
ABL90390  
ID ABL90390 standard; cDNA; 1703 BP.  
XX AC ABL90390;  
XX 24-MAY-2002 (first entry)  
XX Human polynucleotide SEQ ID NO 952.  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; human; secreted protein; gene; ss.  
XX Homo sapiens.  
XX WO200190304-A2.  
XX 29-NOV-2001.  
XX 18-MAY-2001; 2001WO-US016450.  
XX 19-MAY-2000; 2000US-0205515P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Birse CE, Rosen CA;  
XX WPI; 2002-122018/16.  
XX P-PSDB; ABB89981.  
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
XX prevention of neural, immune system, muscular, reproductive,  
XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative

disorders.

Claim 4; SEQ ID NO 952; 2081pp + Sequence Listing; English.

The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB89040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)

Sequence 1703 BP; 324 A; 569 C; 436 G; 364 T; 0 U; 10 Other;

Query Match 3.6%; Score 18; DB 6; Length 1703;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 CCACCTTACCACCACTGG 301  
|||  
DB 1063 CCACCTTACCACCACTGG 1080

RESULT 45  
AAZ24887

ID RAZ24887 standard; DNA; 1715 BP.

AC AAZ24887;

XX

XX

DT 02-DEC-1999 (first entry)

XX

DE Human secreted protein gene 77 clone HPXKT05.

XX

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

XX

XX WO9947540-A1.

XX

PD 23-SEP-1999.

XX

PF 18-MAR-1999; 99WO-US005804.

XX

PR 19-MAR-1998; 98US-0078563P.  
PR 19-MAR-1998; 98US-0078566P.  
PR 19-MAR-1998; 98US-0078573P.  
PR 19-MAR-1998; 98US-0078574P.  
PR 19-MAR-1998; 98US-0078576P.  
PR 19-MAR-1998; 98US-0078577P.  
PR 19-MAR-1998; 98US-0078578P.  
PR 19-MAR-1998; 98US-0078579P.  
PR 19-MAR-1998; 98US-0078581P.  
PR 01-APR-1998; 98US-0080312P.  
PR 01-APR-1998; 98US-0080313P.  
PR 01-APR-1998; 98US-0080314P.

XX

(HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;  
XX Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;  
PI Shi Y, Moore PA;  
XX WPI: 1999-562050/47.  
DR P-PSDB; AAY41384.  
XX New isolated human genes, useful for diagnosis and treatment of e.g.  
PT cancers, neurological disorders, immune diseases, inflammation or blood  
disorders.  
PS Claim 1; Page 344-345; 484pp; English.  
XX This sequence represents a nucleic acid molecule which encodes a secreted  
human protein. The gene number, and the clone it is derived from, are  
detailed in the descriptor line. The gene can be used to generate fusion  
proteins by linking to the gene to a human immunoglobulin Fc portion  
(e.g. AAZ24802) for increasing the stability of the fused protein as  
compared to the human protein only. The invention relates to 95 novel  
genes and their fragments (nucleic acid sequences: AAZ24811-Z24907; amino  
acid sequences AAY41308-Y41404) which are useful for preventing, treating  
or ameliorating medical conditions e.g. by protein or gene therapy. Also,  
CC pathological conditions can be diagnosed by determining the amount of the  
new polypeptides in a sample or by determining the presence of mutations  
in the new polynucleotides. Specific uses are described for each of the  
95 polynucleotides, based on which tissues they are most highly expressed  
in (see AAZ24811 for described uses)

XX

SQ Sequence 1715 BP; 356 A; 557 C; 443 G; 359 T; 0 U; 0 Other;

Query Match 3.6%; Score 18; DB 2; Length 1715;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 CCACCTTACCACCACTGG 301  
|||  
DB 950 CCACCTTACCACCACTGG 967

Search completed: June 16, 2004, 17:43:50  
Job time : 252 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:02:16 ; Search time 2074.33 Seconds  
(without alignments)  
10468.356 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934  
Perfect score: 501  
Sequence: 1 tgcgttcgctgctgcgcct.....cgctgtgctcggtcggtg 501

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 15

Total number of hits satisfying chosen parameters: 9189

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sv.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	501	100.0	53784	1	AMM223012 Amycolato
2	501	100.0	53789	6	A69720 Amycolato 3
3	501	100.0	90445	1	AF040570 Amycolato
4	47	9.4	3900	1	AF262754 Amycolato
5	41	8.2	58343	1	AB089954 Micromono
6	41	8.2	132544	1	AF521085 Streptomy
7	36	7.2	20394	1	AF132222 Streptomy
8	36	7.2	20394	6	AX067996 Sequence
9	36	7.2	43280	1	U78289 Streptomyc
c 10	36	7.2	84985	1	AJ278573 Streptomy
c 11	33	6.6	30000	6	AX250263 Sequence
c 12	33	6.6	103450	1	AF440781 Streptomy
13	32	6.4	20235	1	M63677 S. erythrae
14	32	6.4	20235	6	AR049368 Sequence
15	32	6.4	20235	6	AR095529 Sequence
16	32	6.4	20444	1	SRERYAB
17	30	6.0	1209	1	AF283054 Saccharop
c 18	29	5.8	77534	1	AF235504 Streptomy
c 19	29	5.8	77536	6	BD235937 Polyketid
c 20	29	5.8	77536	6	AR271638 Sequence
21	29	5.8	78210	1	AB070949 Streptomy
22	29	5.8	113193	1	AF357202 Streptomy
23	29	5.8	113193	6	AX703543 Sequence
c 24	29	5.8	138203	1	AY310323 Streptomy
c 25	29	5.8	300425	1	AP005022 Streptomy
c 26	28	5.6	82746	1	AF453501 Actinosyn
27	27	5.4	665	6	AX153707 Sequence
28	27	5.4	15872	6	AR362090 Sequence
29	27	5.4	15872	6	AR362091 Sequence
30	27	5.4	15872	6	BD217171 DNA encod
31	27	5.4	28732	1	AF082100 Streptomy
32	27	5.4	74787	1	AF580915 Streptomy
33	27	5.4	104326	1	SPAS80915 Streptomy
c 34	27	5.4	320150	1	AB070940 Streptomy
35	26	5.2	17101	1	AF005033 Streptomy
36	26	5.2	41097	1	AB017641 Micromono
37	26	5.2	123580	1	AF016585 Streptomy
38	26	5.2	125401	6	AF263912 Streptomy
c 39	26	5.2	132544	1	AX211739 Sequence
40	25	5.0	9769	1	AF521085 Streptomy
41	25	5.0	24568	1	SEERYA S. erythrae
42	25	5.0	50937	6	AF220951 Streptomy
43	25	5.0	76196	1	AF159871 Sequence
44	24	4.8	3978	6	AY354515 Streptomy
45	24	4.8	3978	6	BD269390 Mitomycin
					AR266876 Sequence

ALIGNMENTS

RESULT 1  
AMM223012  
LOCUS Amycolatopsis mediterranei genes encoding rifamycin polyketide  
DEFINITION syntheses, ORFs 1 to 5.  
ACCESSION AJ223012  
VERSION AJ223012.1 GI:2764760  
KEYWORDS ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.  
SOURCE Amycolatopsis mediterranei  
ORGANISM Amycolatopsis mediterranei  
Bacteria; Actinobacteridae; Actinobacteriales; Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.  
REFERENCE 1  
AUTHORS Schupp,T., Toupet,C., Engel,N. and Goff,S.  
TITLE Cloning and sequence analysis of the putative rifamycin polyketide

AMM223012 53784 bp DNA linear BCT 09-FEB-1998  
Amycolatopsis mediterranei genes encoding rifamycin polyketide  
syntheses, ORFs 1 to 5.

```

synthase gene cluster from Amycolatopsis mediterranei
Unpublished
REFERENCE
2 (bases 1 to 53784)
AUTHORS
Schupp, T.
TITLE
Direct Submission
JOURNAL
Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma
AG, CH 4002 Basel, Postfach, Novartis K-681.344, 4002 Basel,
SWITZERLAND
FEATURES
Location/Qualifiers
1..53784
/organism="Amycolatopsis mediterranei"
/mol_type="genomic DNA"
/strain="LBG A3136"
/db_xref="taxon:33910"
1336..15543
/function="CoA ligase, acyltransferase, ketoreductase"
b-ketocacyl-ACP-synthase, acyltransferase, ketoreductase"
/notes="ORF1
PKS modules 1-3"
/codon_start=1
/transl_table=11
/product="rifamycin polyketide synthase, type 1"
/protein_id="CA11035.1"
/db_xref="GI:2764761"
/db_xref="GOA:O54666"
/db_xref="SPTREMBL:O54666"
/translacion="MRTDLIKPLHVLLENATRFAGKPAFADHRTVTYTGDLARTTR
LAGHLAGLVGHGRDRAICLGNRVSTVESYFALLRAGAGVGPLNPGSATALEHPLTRD
SGATVVTDAAQARLARLAEHVELLVTDGDDVPEGAHSDYDELASEPAPADDELDE
PAMFYTSTGTPKGVVSTORNCIMSVASCYTFPGGLVDQDQDRLWPLPLFSLHSHIA
CWSVTSVARIADGSSADVWMLTEASSTVPLAGVPTTYHHLVRAARQGRFSAPS
LRGLAGAGVGLRSEFETFCGLIDAYGSTETCGCAITMNPDDGARVEGSGCLAV
PGVDRVVDPTGLDPAGEGEVWVGNPMLGYHNSPRATAAMDDGFRGTGLDAR
RDAGYPTICGRILKELIRGCANTHPCGEVAVLRTDVGADAAVGVPHDTLGEVPA
VYIPGTPGPDAALIEKCRQLSAYKVPDRILEVAHLPRTASGKIRKGLLTDPAQUR
YAATEHEQRSHADESVAARLARKLSGLDERAQCELEDLVRTQAADVLTQDPPQRA
FRDGTGSLAIVELNRKLTHETGLWLPASAVFDPHTPAALAAARVRAEELGITQAVAP
VVAADGPEIATVGMACRLPGVASPEDLWRLVAERPDVAASEFPGDRGMDLSLIDPD
RERAGTSGVGGGLHDAGFDAGFCGISPREAVAMPQOORLLLETSEWALENAGVDP
IALKGTDTGVFGMLGGYGSGVAPALEGFTVTGVASSVASGRVSVLGLEGPATV
DTACSSSLVAMHLAAQALROGECSMALAGVTWATPGSFVSFRQALAPDGRCKAF
AAADGTGSGEGVSVVLERLSVAREGRHLILAVIRGSAVNQDGSNGLTAPNLSQO
RVTRRAALAGLAPDGVVVEAGTGTGLDPTLAEQALLATYGOERKQPLWGLSKSN
IRYRRAVDYASHTHRVEDIRDTLAETLAGI SAQAPAVPFYSTVSEWRDAGVLDGG
WYRLNRQNVFGAAALALEOQHTFVESAHPVTQPLSELGTDAIGTLRREDGGL
RLLASMGELFVRGIDVDMTAMVPAAGWDLPTVAFRHHVWLEPAPASAGDPLICT
VUSTPGSDELTAQAQSERAPQWAVDGLVENAALVEAAILRLGLAGTFFVUGELVDP
VUUPRGSREVQLIIGBPGCRHPTIEFSEADPEWIRHAGHTLAPAAAVPEPAAA
GDATVTVAGLRDADRGHHPALLDAARTVVGDDLLPSVMTGSLASGATATVTP
TATGLRITDPAGQPLTVESVRGTFVFAEGCTDALARVDWPELPTLPTAETDPLPYE
ATGAETLSALQAWLADPAETRLAVTGDCTPCGAAILWGLVRSQAQSEHCPRIVLADL
DDPAVLPAVASGEPQVRVNGVASVPRLTRETPQDARPLDPEGTVLITGGTGLCA
LTPARHLTAHGHVHLVLSRGEAPELOEBELTLAGASVAIACDVADRAQLEAVLRAI
PAEHLPTAVLHITAGVLDGDDVVTEPDRLATVRPKVDAAKLLDREADLAAFLVLP
SSAQRNGVCGQAGAAANAEALDALQRNSLDLPAYSIAWGYWATVSGMTEHLGDAD
LRRNRQICMSGLPADEGMALIDAAIATGGTILVAKFDPVAAALRAKAGGVPPTLRLG
APLPRAAAKTASLTLERLAGLAETEQAALLDVRHEAAVLEHSGASVHRSPFTKO
AGPDSLTAVELNRLLAATATGILTSPMIEDYKPKPALADHLRAKFGSAANRPAEICT
AAAEPTAIVAMACRPFPGVHSPEDLWRLVADGADATEPFDRCNWTDRLYHEDDPH
EGTITYRHGAFDDDAADFAGGISNEALAMPDQOORLLLETSEWELFERAADIPTTL
AQDIDYVFGVNSHDYSNMRHRAAGVEGFRITLGGASVLSGRVAFHVGVEPAAVTT
ACSSSLVAMHLVQALORGECSMALAGVWMTGVTFVFSRQRLAGDPGRCKAFAD
GADGTGSEGVGLLQLERLSEAQRHQHVLAVVRGSAVNSDGNGLTAPNPGSQOVR
IRKALAAGTSDVDAVEAGTCTTGLDPTIEALLATYQNEETPLWLSVKSNLG
HQAQAAGVAGVIKMAMRHGVLPRTHLVDRPSSYVDWSAGAVELLTEARDVSNHGA
RRQVSGFSGGINNAHVLEVAAPITTPQEPAPFELVPLVLSVATAAGLRQGRHIA
AFLGDRITVRVPDAAVALATRAQLDHRVAVLASDRAQLCADLAAFGSGVVTGTFVDG

```

## CDS

```

KLAVFTGQSGWAGMGRLEARTFPVFRDAFEACEAVDTHLRERPLREVVFDSDALL
DOTMYTCGALFAYETALFRLFSWGRVPLGLAGHSIGELAAAHVSGVLDLADAGELVA
ARGLMQALPAGGAMVAVQATEDEVALPDDTVCVAAVNGPDSVVLSTRAAVALIAD
ELAGRKTRERLAVSHAFSPLEMPEDLDPRVAERLTYPAGSLPVYSTTGLGELAALD
SPDYWGQVNRVRFSDAVTALGAQAGSTFELGPGGALAAALGTTGGPQSCVATL
RKNGAEVDPDLTALAEHLVHGVDWMTTTLDEPATAVGTVLTPTFAFOHQFSDVDEET
AAVSVTPPPAAPTIVDRPVQDRELRESAAVILGHRDAGSFDLDRSEKDFGHSFSLAV
KMLRLRDLFTGVLPSTLIIDYPNPAVLADHLRAELIGERPAAPAPTRVDSDEPIAI
VGMSTRLPGAGDSPEELKLVAGRDVAGSFPVDRGMDLGLVHPDPHAGTSTVRGG
GFLHDAQDFDAGLFGISPREALAMPQOORLLLETSEALERAGVDPPLSARSGSDVGT
GIYHHDYVTKLRVEDPVQGTWTGTATSSVASGRVAVTFGEFGPAVTDATACSSSLVA
MHIAAQALROGECSMALAGATVWASDALEFRSQRGLSADRCCKAYASAGDGTGWA
EGYGVVLERLSVAREGRHVLAVIRGSAVNQDGSNGLTAPNPGSQOVRVIRGALASA
GLAPSDVVDVEGHGTGTALGDPTEVQALLATYGOERQPLWGLSKNLGHTQAAAGV
GVGIMAMRHGVMPTATLHVDRBRTSVDMSAGAEVLTBAREWPTGRPRAGVSF
GASGTWHLIIIESGPAEEADEVESVTVPLVVSARSAGSLAQOGLAAVLENESLAG
VAGALVSGRATLNERAVVIAGSDEDAQOGLQALARGENAPGVVTGTAGKGVVWVFP
GQSQWMGMDRLDSSPFAARIKECAALEQWTDWLSLDDLRDADLLDRVDVQF
ASFAMVGLAAVMTSLGVTDPDLVHGSQGEIAAACVSGALSDDAAKVAALRSQAIAG
ELAGRGMASVALSEEDAVARLTPTWNRVVEVAAVNSPVSIVIAGDAQALDEALELAG
DGVRRVAVDYASHTHREAIATLAKTLAGIDARVPAIPFYSTVLTGITEQAVDVA
GYMYRLRQOVRGSPVADLAGLGHVTFVEISAPHLVQPLSEISDDAVTVLSRRDD
GGLRLLSAELHYVRGVAVDTAAPAAGWDLPTYAPRRRHFHHEATRAEAGEM
DGEFTWAEQSDVDSLAELLELPQEGALSTVVPVLAQRDRRREKSTAEKURQVT
WQPLERAGVPGGRMLAVVPAAGTTDALLKELTQOGLDITVLEIEASRAQLAQOLRN
VLAEHDLTGVLSLALDGPADAAEITASTLALVQALGDTTTSAPLCLTSGAVNIGI
QDAVTPAQAAVWMLGRVALERLDRWGLVDLPAAIDARTAOALLGVLNCAAGEDOL
AVRSGVYRRLVRKVPESAITSRWRPGTVLTGGAEGILGRHASVLAQSGAERLLIV
TGTGDVDELTAELAEFGTTVEFCADTDRDAIQLVADSEVTAVVHAADIQTSSVDIT
GVADLDFEAKVTTAVMLDQDFEDTDAFVVFSSITAGIWWGGGQPGAGANAVLDA
LWEMRARGALQTSIAMGALDQIGIMDEAALQSLRRRGTVPWAPPLPAVMOQVAG
NEKAVADAMWAAFPATFSVRPSPLFADLPKAKILRAAQDGDGDEGTASSLADSL
KAYEDAEONRILLKVRGHASTVLSHSGAEGIPROAFQVQFSDSLAANLRSLHAA
TGLRLPATLFDYPTPEALVGLRVELREADDGLGREDLREVLAAVPAEPAREKAG
VLDTLGLATGTGHEPDTAETTAAPAAADDAELIDLDSIGLVQALGQTS"
15550..30759
/function="b-ketoacyl-ACP-synthase, acyltransferase,
dehydratase, ketoreductase, acylcarrier protein"
/notes="ORF2
PKS modules 4-6"
/codon_start=1
/transl_table=11
/product="rifamycin polyketide synthase, type 1"
/protein_id="CA111036.1"
/db_xref="GI:2764762"
/db_xref="GOA:O52789"
/db_xref="SPTREMBL:O52789"
/translacion="MANQSWRKWMSAPNEQIVDALRASLKENVRLQENSALAAAAA
PVAIVSMACRYAGIRGEPEDFWRVSEGADVYTGFPEDRGMDEGLXHPDPNPGTTY
VRBGAFLQDAAQDAGFCGISPREALAMPQOORLLLEVSWETLERAGIDPHSVRGSDI
GVYAVGVODYAPDLGSGFMSLERALGTAGGVASGRVATLLEGPVATVDTWCS
SLVAIHLAAQALRGECSMALAGSTWATPGFVGPARORATLADFCRCKSYAAADG
SWAEGVGVLLERLSVAREGRHVLAVIRGSAVNQDGSNGLTAPNPGAQOVRIRKA
LAAGAPSDVTDVTEGHGTUTVLGDPTEVQALLATYQOGRDQOPLMGLSVSVVGH
QAASGAVAGTVMKHLHQEPATQHDVAPTQVDVMSAGAILLEAGWEPNRPGRHRR
GGISFGAGSTNAHMLIEEAPDEPEPTAPAGTVPLVVSAAATAASAAQAGLAEV
GGLSVADAGTGLVSMGRAMLSERAVVAGSHHEATVGLRALARGESAPGLLSGRSGV
KGVWVFPQGGTQWAGMGRLLDSSSEVFAARIACETALGRWDMSTLTVLRGADILL
DRVDVQPASFAVWGLAAVWASLGVPEAVVGHSGEIAAACVSGALSLEDAKVA
LRSAQALASLAGRGMSVALSEEDATAPWAPWAGRVAAVNGTSPVITAGDREADL
EALDLDQDQVIRRVAVDYASHTRHVEAARDALAEMLGGIQAQAPVPPVTGTM
VEDAGLDGQVYTRRQVRFQGPVAAELTQOHRVFEVESAHPVLVQPTINELVDDTE
AVVTGLRREDGGLRLLSAELVVRGTVDMNSGVLPPRRRVELPTPAGTGLVMSREVSAAQ
MGSATDASLGLAGADHPLLAGAVPLQSDGLVFTSRLSLKSHPLMAGHAIIGVLLI
PGTVVDLALRAGDELGFGLVEHLIEAPLVLGERGVRVQVAVSGNETGSRADVDF
SMEDGDWETWTHLGLIGASTSEPSRDFFAAPPAGEPIDVENFTYDITERGAYTS
QKQGRVARRRGDEVFVALEPDDREDKAGTGLHPALLDAAHNTAFNAPDDDRSV
LPTANGLVILVHAGSALRVVAPGGPDALTFOAADTGGVLTMSLVREVSAAQL
ETAAGERDLSIFOVDMTEPATTAETAEVLEAFCEAAPELTSRLBAVOSWLAD
AADARLVVTVTRGAVREVTDPAGAVKGLVRAQAENPGRIILVDITDGDVPLGAVLAS
GEFQLAVRGNAFSVPRIARATGEVPEAPAVFSPEGTVLLTGGTSGGLVAKHLVARH
GVRRLVLASRGSVAENEDLVTELTPEQPAVNSVACDSDQDQVAAALLAEABRGTIVLHIA
GLLDDGTVIGALNRERLAGVPKVDVAVQHLDELTRDILGLDAFVVVFSSAAALMSGAGQ

```

[illegible]

Query Match	100.0%;	Score 501;	DB 1;	Length 53784;
Best Local Similarity	100.0%;	Prod. No. 9.6e-228;		
Matches 501;	Conservative	0;	Mismatches 0;	Indels 0;
Gaps	0;			
1	TCGTCCTTCGCTGGTCGCGCTGCACATGGCGCGTGCAGGCCCTGCAGCGCGGCGAGTGCCTCC	60		
8434	TGCTCTTTGCTGGTTCGCGCTGCACATGGCGCGTGCAGGCCCTGCAGCGCGGCGAGTGCCTCC	8493		
61	ATGGCGCTCGCGGGCGGGCGGTGATGGTGTGATGGGCACGGTTCGAGACGTTTCGTGCGAGTTCTCG	120		
8494	ATGGCGCTCGCGGGCGGGCGGTGATGGTGTGATGGGCACGGTTCGAGACGTTTCGTGCGAGTTCTCG	8553		
121	CGGCGAGCGGGCTGCGCCCCCGACGGCCCGCTGCAGAGCGGTTTCGCCGAGCGGCGGACGGC	180		
8554	CGGCGAGCGGGCTGCGCCCCCGACGGCCCGCTGCAGAGCGGTTTCGCCGAGCGGCGGACGGC	8613		
181	ACCGGCTGTCGCGAGGGGCTCGGGCTGCTCTCTGGTGGAGCGGCTGTCCGAGGCTCAGCGT	240		
8614	ACCGGCTGTCGCGAGGGGCTCGGGCTGCTCTCTGGTGGAGCGGCTGTCCGAGGCTCAGCGT	8673		
241	CGGGGGCACAGGTCTTGCCTGCTGCTCGCGGGTTCGGCGGTCGAATCGGACGGCGCTCG	300		
8674	CGCGGGCACAGGTCTTGCCTGCTGCTCGCGGGTTCGGCGGTCGAATCGGACGGCGCTCG	8733		
301	AACGGCTTCAGCGGCCCGCGAAGCGGCCGCTCCAGCAGCGCGTGTATCCGCAAGGCACTGGCC	360		
8734	AACGGCTTCAGCGGCCCGCGAAGCGGCCGCTCCAGCAGCGCGTGTATCCGCAAGGCACTGGCC	8793		
361	GGCGCGGACTGTCCACATTCGGAAGCTTCGACGCGGTGGAGGCGGCACGGCACCGGACGACC	420		
8794	GGCGCGGACTGTCCACATTCGGAAGCTTCGACGCGGTGGAGGCGGCACGGCACCGGACGACC	8853		

QY	421	CTGGGCGAACCCGATCGAGGCCGAGGCGCTGCTGGCCACTTACGGCCAGAAACCGGGAACG	480
Db	8854	CTGGGCGAACCCGATCGAGGCCGAGGCGCTGCTGGCCACTTACGGCCAGAAACCGGGAACG	8913
QY	481	CCGCTGTGGCTCGGTCGGTG	501
Db	8914	CCGCTGTGGCTCGGTCGGTG	8934
RESULT 2			
LOCUS	A69720	Sequence 3 from Patent WO9807868.	
DEFINITION	A69720	Sequence 3 from Patent WO9807868.	
ACCESSION	A69720	Sequence 3 from Patent WO9807868.	
VERSION	A69720.1	GI:4774328	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE	1	(bases 1 to 53789)	
AUTHORS	Schupp, T., Toupet, C. and Engel, N.		
TITLE	RIFAMYCIN BIOSYNTHESIS GENE CLUSTER		
JOURNAL	Patent: WO 9807868-A 3 26-FEB-1998;		
FEATURES	CIBA GEIGY AG (CH)		
source	Location/Qualifiers		
	1..53789		
	/organism="unidentified"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:32644"		
ORIGIN			
Query Match	100.0%;	Score 501; DB 6; Length 53789;	
Best Local Similarity	100.0%;	Pred. No. 9.6e-228;	
Matches 501; Conservative	0; Mismatches	0; Indels	0; Gaps
QY	1	TCGCTTCGCTGGTTCGGCTGCACATGCGCGGTGCAGCCCTTCGACGCGCGGCGAGTGTCC	60
Db	8434	TCGCTTCGCTGGTTCGGCTGCACATGCGCGGTGCAGCCCTTCGACGCGCGGCGAGTGTCC	8493
QY	61	ATGGCGCTTCGGGGCGCGTGGTGGTATCGGCGACGCTCGAGAGGTTCTCGAGTTCCTCG	120
Db	8494	ATGGCGCTTCGGGGCGCGTGGTGGTATCGGCGACGCTCGAGAGGTTCTCGAGTTCCTCG	8553
QY	121	CGGCGAGCGCGGCTGGCCCGCGCGCTGCAGCGGCTTCGCGCGACGGCGCGGACGGC	180
Db	8554	CGGCGAGCGCGGCTGGCCCGCGCGCTGCAGCGGCTTCGCGCGACGGCGCGGACGGC	8613
QY	181	ACCGCTGGTTCGAGGCGCTCGGCTCTCTCTGTGTGAGCGGCTTCGAGGCTCAGCGT	240
Db	8614	ACCGCTGGTTCGAGGCGCTCGGCTCTCTCTGTGTGAGCGGCTTCGAGGCTCAGCGT	8673
QY	241	CGCGGGACACAGTCTTCGCGCGTGGTTCGCGGGTTCGCGGGTCACTTCGAGCGGCGCTCG	300
Db	8674	CGCGGGACACAGTCTTCGCGCGTGGTTCGCGGGTTCGCGGGTCACTTCGAGCGGCGCTCG	8733
QY	301	AACGGCTTGACGGCCCGCGCGCGCTCCAGCAGCGCGTGTATCCGCAAGGCACTGGCC	360
Db	8734	AACGGCTTGACGGCCCGCGCGCGCTCCAGCAGCGCGTGTATCCGCAAGGCACTGGCC	8793
QY	361	CGCGCGGACTGTCCACATCGACAGTTCGACGCGTGGAGCGCACGCGACCGGACGACC	420
Db	8794	CGCGCGGACTGTCCACATCGACAGTTCGACGCGTGGAGCGCACGCGACCGGACGACC	8853
QY	421	CTGGGCGAACCCGATCGAGGCCGAGGCGCTGTGGCCCACTTACGGCCAGAAACCGGGAACG	480
Db	8854	CTGGGCGAACCCGATCGAGGCCGAGGCGCTGTGGCCCACTTACGGCCAGAAACCGGGAACG	8913
QY	481	CCGCTGTGGCTCGGTCGGTG	501
Db	8914	CCGCTGTGGCTCGGTCGGTG	8934



```
RESULT 3
AF040570 90445 bp DNA linear BCT 16-APR-2001
LOCUS
DEFINITION
ACCESSION AF040570 AF040571
VERSION
KEYWORDS
SOURCE
ORGANISM
Amvicolatopsis mediterranei
Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
Pseudonocardiaceae; Pseudonocardiaceae; Amycolatopsis.
REFERENCE
1 (bases 1 to 90445)
AUTHORS
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R., Yu, T.W.,
Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X., Hutchinson, C.R. and
Floss, H.G.
TITLE
Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
the molecular analysis of the rif biosynthetic gene cluster of
Amvicolatopsis mediterranei S699
Chem. Biol. 5 (2), 69-79 (1998)
JOURNAL
MEDLINE
PUBMED
98174059
9512878
REFERENCE
2 (bases 1 to 90445)
AUTHORS
Kim, C.G., Yu, T.W., Fryhle, C.B., Handa, S. and Floss, H.G.
TITLE
3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
formation of the precursor of mC7N units in rifamycin and related
antibiotics
J. Biol. Chem. 273 (11), 6030-6040 (1998)
JOURNAL
MEDLINE
PUBMED
98165773
9497318
REFERENCE
3 (bases 1 to 90445)
AUTHORS
Yu, T.W., Muller, R., Muller, M., Zhang, X., Draeger, G., Kim, C.G.,
Leistner, E. and Floss, H.G.
TITLE
Mutational analysis and reconstituted expression of the
biosynthetic genes involved in the formation of
3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin
biosynthesis in amvicolatopsis mediterranei S699
J. Biol. Chem. 276 (16), 12546-12555 (2001)
JOURNAL
MEDLINE
PUBMED
21201076
11278540
REFERENCE
4 (bases 1 to 90445)
AUTHORS
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R.,
Hutchinson, C.R. and Floss, H.G.
TITLE
Direct Submission
Submitted (30-DEC-1997) Dept. of Chemistry, University of
Washington, Box 351700, Seattle, WA 98195-1700, USA
JOURNAL
REFERENCE
5 (bases 1 to 90445)
AUTHORS
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R.,
Hutchinson, C.R. and Floss, H.G.
TITLE
Direct Submission
Submitted (19-JAN-2001) Dept. of Chemistry, University of
Washington, Box 351700, Seattle, WA 98195-1700, USA
REMARK
Sequence update by submitter
COMMENT
On or before Jan 22, 2001 this sequence version replaced
gi:2792340, gi:2832263.
FEATURES
Location/Qualifiers
1..90445
/organism="Amvicolatopsis mediterranei"
/mol_type="genomic DNA"
/strain="S699"
/db_xref="taxon:33910"
423..1451
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC01706.1"
/db_xref="GI:2792310"
/transl_table=11
/translation="MKTGGVGSRYVYKTLRQLEVALIGAGLIRLHLFWLGGAAV
RVYSDGRSRELAEEFGAKAAGSLBEALDGDADVDICTPTASHEITALTAAGVGW
CEKPLAASAEAEIYTAERAGRVLYAAHDVFAAPYARLHELVSAGRLGEGALGRF
SFAPYAPRWTHASARSQGLLDQLHGLADLHNVFGDVRVHACYQGDIAPEEG
AVATGVATHASGAI SQVSVRWTAIPRPVRFVAFVHSGTGSVSDSEMPQVRVVD
GGAGNFAYGGGVSFVDEMFATAFAGGPEPRIGAKDALAAVRIIHAASAEAWTGRAV
ELPVGAA"
```

```
1448..2215
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC01707.1"
/db_xref="GI:2792311"
/translation="MKVAILSSSTPQAYAGALRGLDVEVVAASWDAFEFVRQAEAG
ARVLCYPPAAKETDLMKAMIDAAAGRLTFASPAACHGFAFVNRIGIAGGIGELTTL
SGVATSDVGLGAAAPYLLDLADAVLGGEPAAQVYAQTNTVLSRIGESAAVLTFRYR
SGVASPDCHRHGSAATGLPAVTFTGDQGSQVYDAGPOLGGERPELGGEDLEALMKD
FLAGDGPDPGQAALRTFLIIQAAYESAHTGQPVDL"
2353..2619
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC01708.1"
/db_xref="GI:2792312"
/translation="MQDRSVLGGDPKQKQVEAMSVTKIALSGGPAELAPAHRSVDSG
ALENTLKIRHGAGYEHFVHGGEFQAVDCCDIAPFRWSHRTKIAE"
2652..3842
/note="ORF0; similar to Streptomyces Sp. Chop"
/codon_start=1
/transl_table=11
/product="cytochrome P450 monooxygenase"
/protein_id="AAC01709.1"
/db_xref="GI:2792313"
/translation="MTDAISPEVPWDRDKFDPDPVAFDLSREERPLAKMVPDGHVGW
IVSSVELRVLSLDRFSHSCVGHFPVTHQGVPTPHLPFGMFIHMDPEHLYRK
LITGEFVRRASRLIPRAEVAABQIEVMRAKAPADVMDFAKPLVRLMGLVGLP
FEEDRIVPAVILLHDADPAEAAAAYEYAKGFDEVIERRORPODDLISLVTED
LTQBELRNVTLLFAGYETGALATYFALLHHTDQALRALEPEKLDAAIEELLR
YLVNQYHTALEVDKLEGLIKKGTDTVTSIPAANRDPKAGCPAELDIEDTISG
HVAFGFIHQCLGQMLARIELRAGFTALLRAPPFLRAVPADAEVFLRLKGSVFSYKKL
PVSW"
3883..18090
/feature="rifa"
3883..18090
/feature="rifa"
/feature="rifa"
/feature="rifa"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAC01710.1"
/db_xref="GI:2792314"
/translation="MRTDLIKPLHVALLENATRFAGKPAFADDDHRTVTVDLEARTR
LAGHLAGVGRHGRDVAICLGNRVSTVESYFALIRAGAVGVPLNPGSATALEHPLTD
SGATVVTDAQAARLRLAPHLVAVTGDVPEGAHSDYDELASPAEPAADLELDE
FAMWFTSGTGRPGVSTQKNCNLSVASCYVFPFGLSDQDRVLPFLFHSLSHIA
CVLSATVVGASVRIADGSDADDVRLIEASTFFLAGVFTTTHHLVRAARQGFSAFS
LRIGLAGGAVLGLRSEFEETFGVPLIDAYGSTETCGAITMNPDPGARVSGSGLAV
PGDVRVDPDTGLDPAGEGEVWVGNVLMGYHNSPEATAAAMRDGWFRTGLAR
RNDAGVFTICGRIKELIRGGANIHGEVEAVLTVGVADAAGVGHDTLTGEVPA
YVTPGTGFDPAALIEKCEOLSAVKVDPDRIEVAHLPRTASGKIRRGTLTDEPAOLR
YAATHEEQSHADESVAALRSLGDLERAQCELELDVLTQADVLGQVPOGARA
FRDLGTSIAIVELNRNLTGMLPASAVIDHPTPAALAAARVRAELIGITQAVAEF
VVAADGEPFAIVGMACRIPGGVASPEDEMLRLVAERDVADSEFFPGRDGLSDLDPP
RRAGTSYVQGGFLHDAGEFDAGFGISPREAVAMDPOORLLETSEWDRDAGLVDP
IALKTDGTGVSGLMGQGVGSAVAPELEGVTTGTGVASSVSGSVYVGLGEPVTV
DTACSSSLVAMHLAQLRQGSMAAGVTVMTAGTSFVSFRORALAPGCKAF
AAADGTGWSGVGVTVLERLSVAREGRHILAVLGSVAVNODGNGSLTAPNGLSQQ
RVIRALAAAGLAPSDVDVVEAHGTGTTLGDEPLAQALLATYQGERKQILNGLSKEN
IGHAQAAGAVGIVKMQALRHETPLPTLHVDKPTLEVDWSAGAILLEFARAPNG
RPRAGVSFGVSGTNAHLILEEAPAEFPAAPPELVPLVSVASTESLSQAERLA
SLLEGDSLTVEAGALVSRVAVVAGSREAEVTLRALTAGSTGPKVVMV
FPGQGTQWAGMCELLAESPEAEIAECAALAPWDMSLVLDVLRGEGDLGRVDLQ
PACFVAGLAAWESVGRVPAUVHSGCEIAAACVSGALSIEDAAKVVALRSOAI
AELSGRGNSVALGDDVVSRVLDGVEVAANGFSSVVIAGDAHALDALTLEILSEG
IKVRVYVASHTRVEDIRDTLAETLAGISAQAPVFFYTSVTSEWDRDAGLVDP
YVNRNVRNVRGAATALLAQCHTVFVEVSAPVTVQPSLTGAIETLRDGL
RILLASMGELFVRGIDVMTVPAAGWDVLTPTAFERHRYWLEPAEPASAGPLLGT
VTPGSDRLTAVACSRRAQWADGLVPAALVEAATRLGLACTPVVGLVADAP
VLPKRGSEVQLI VCEPGEQRRIEVSREADEPWTTHAGTTLAPAAAAPPEPAA
GDATDVTVAGLDADRYGHPALLDAAVTVTVGDDLLPLPSVWTSGLASGATVTVTP
```



TATGTLTDPAGCPVLTVESVRGTPFVAEQGTGTTDALFRVDMPEIPLPTAETADFPLPYE  
ATSAETLSALQWLADPAETRLAVTGDCTEPGAAATWGLVRSQASHPGRIRI VIADL  
DDPAVLPAVASGEPQVRNRGVASVPLRTVTTPQDARPLDPEGTVLITGSGTGLGA  
LTLARLHTAHGVHVLVSRGEBELQELTALRGASVAIAACDVADRQAADLAARLAI  
PAEHLPLTAVTHAGTVDGVTVELTPDRLATVRPKVDAARLLDELITRAQLAAVAF  
SSAAGVINGGQGAAYANAELDAQRNSLDLPAVSATWGYATVSGMTEHLGDAD  
LFRNRIGIMSGLPADEGMALDDAATATGCTLVAAKFDVAALRATKAGGVPPLRLGL  
APLPRAAKTASLTRELGLAEIPQAAALLDVRHAAVLHSGASVSHSPRFTKO  
AGFDSITAVELNRLAAATGLTSPALPDYKPKPALADHLRAKLPFGSAAANPRTGT  
AAAEBFIALVAMACHPFGCVHSPEDLWRLVADGADAVTEFPADRGWDITRLIHEBDDH  
EGTTVRHGAFDDDAAGFDAAFFGSPNEALAMDPOORLLLETSELWELFERAIDPTTL  
ACQDITGVAGNVNDSXMRHFAAGVEGFRLLTGGASVLSGRVAVHYVGEPAVITVT  
ACGSSILVHMAVQALQRCESMARAGHVVMTVETFEFRQRLGADPDCKAFAD  
GADGTWGBGVGLLVERLSEARQGHVLA VRGSVANSNGASGLTAPNPGSOORV  
IRKALAAAGLSVDVDAVEAHGTITLGDPIEAELATYTGONRETPILWLSGVNGLP  
HTQAAAGVAGVITVAMMRHGVLPRTLHVDRPSSVYDWSAGAVELLTEARDVWSNGLP  
RGAAGVSGIGKTNHVVLEVEAAPIITPQBPFAFLVPLVLSRAGLQAGRLA  
AFLGRTDVRPDAAYALATTRAQLDHRAVLASDAQICADLAAFGSGVVTGTPDG  
KLAVLFTGCGSQWAGRELAETFPVFRDAFEAACEAVDTHLRERPLREVDFDGLL  
DOTMTQGFALFAVEETALPFLFESWVRPCLLAGHSIGELAAHVSGVLLDADAGELVA  
AEGRLMOALPAGAMVAQVATEDEVALPILDGTVCVAVNGPDSVLSGTDEAVLVA  
ELAGRGRTRRLAVSHAFHSPLEMPMLDDFRAVAERLTYRAGSLPVVSTLTGELALD  
SPDYVQVRNARVSDAVTALGAQASTFLFELGGGALAAWALGLTGPGSCVATL  
RNGAEVDPVLTALELHVTRGVGQDWTVDLDEPATVATGTVLPTAFQHOFVWDVDET  
KLNRRLDPTGVELPSTLLFDYPNPAVLADHLRAELLGRPAAPAPVTRDVSDEPIAI  
VGMSTRLPGGADSPELKLVAERDASVGFVDRGMIDLGLYHPDPARAGTSYTRSG  
GLIHHDVTRLREVDPVGYMTGTASSVAGRVAYVFGFEPAPVATVATACSSSLVA  
MELAAQALQRCESMALAGATVMAASPDFAFLFESRQGLSADGRCKAYAEAGDGTGWA  
EGVGVVTLERLSEARQGHVLA VRGSVANSNGASGLTAPNPGSOORVIRGALASA  
GLAPSDVGVHGTGALTGLEVOALLATYGOEREPMLWLSGLKSNLNGHTCAAAGV  
VGVITKIMAMRGWTPALTVDERTSOQVDSAGAIKLVTEAREWPTGRPRAGVSSF  
GASGNTAHLIIIEGPRAEAEVDEVASVPLVVSARSAGSLAQAGLAAVLNESLAG  
VAGALVSGRATLNRARVITAGSRDEAQDGLQALARGENAPVVTGTGPRKVVVVFP  
GGGQMMGRDLSSPFAARIKECAALEQWTDMSLDLVLRGDADLDRVDDVVP  
ASFAMVGLAAVLSIGVTPDVLVHSGOGEIAAACVSGALSDDAKVVALRSQALAG  
ELAGRGMSASVLSDEDAVRLTPWANRVEVANSVSSVWTAGDAQALDEALEALAG  
DGVRVRAVDVASHRSHVEAETLAKTLAGIDARVPAIPFVSTVLGTWISOAVDA  
GQWYRLRQAQVRFGPSVADLAGLHVTVFEISAHVPLVQPLSEISDAVVTLSLRDD  
GGRLNLLAARELIVRGVADWMTAAVPAAGVNDLPTIADFRRHFWLHEATEAEAGD  
GVADLDEVFAAKVTTAVWLDOLPEDTFLDAFVVFSSINGVGGGQGPAGANAVLDA  
LVEMVRRARGLKATSIANGALDQIGIMDEAALAAQLRRRGVTPMAPPLAVTAMVQAVAG  
NEKAVADMDWAAFTIPTSVRSPPLFADLPEAKAILRAAODDGDGTAGSLADSL  
RAVPDAEONRILIKLVHGHASVILHSGAEGIGPROAFOEVGFDLSLAANLRSNLHAA  
TGLRLPATLIPDYTPTEALVGVILVRELLREADDGLDREDLRLVLAANVFPARFKAG  
VLDTLGLADTGTGPTGPTDAETTEAAAPADDAELIDLISGLVQALGQTS"  
/misc\_feature  
3883..5775  
/genes="rifa"  
/notes="encodes polyketide synthase loading domain"

Query Match 100.0%; Score 501; DB 1; Length 90445;

Best Local Similarity 100.0%; Pred. No. 9.1e-228;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCTCTTCGCTGCTCGCGTGCACATGCGGCGTGCAGCGCTCGAGCGCGCGAGTGTCTCC 60  
DB TCGTCTTCGCTGCTCGCGTGCACATGCGGCGTGCAGCGCTCGAGCGCGCGAGTGTCTCC 11040  
QY 61 ATGGCGCTCGCGCGCGCGTGTGATGGGCACCGTTCGAGACGTTTCGTCGAGTCTTCG 120  
DB ATGGCGCTCGCGCGCGCGTGTGATGGGCACCGTTCGAGACGTTTCGTCGAGTCTTCG 11100  
QY 121 CGGACGCGCGGCTGCGCGCGCGCGCGTGTGAGAGCGGCTTCGCGACGCGCGCGAGCGG 180  
DB CGGACGCGCGGCTGCGCGCGCGCGCGTGTGAGAGCGGCTTCGCGACGCGCGCGAGCGG 11160

```

VALALATRAQLDHRVAVLASDRDLRADLAAPGPGVVTGTGTPVDGKLAVLFTGGSGOM
AGMCRELAGTFPVNRAFEACHADHLAGRAVREVVFDTRDLDTQMTVOGALFAV
ETALFLRFASNGVRPDLLAGHSITGELAAHVUSGLDLAEAGELVAAAGRLMOALPAGG
AMVAVQAEDEAVPLDGTVCVAVANGPDSVLSGAEDAVLAANAAGLAARGRKTKRLP
VSHAFSPLEMEPLMDGFPRAVABRITYRAGTLFVSTLTGLKVALADTGTGVVVRQVRDAV
RFGDAVTNAGATTYLELGGGALAAALGTGGPEQSCVATLRKXGAEVPGVLTA
LAELHVRVQWVTWALDRPGTAVGTVLPTVAFQHERYVWGTDTSPSPASPVAAAPAE
PVHDVPAVAVLGHRTDAPDLDRSKDGLGFSI GAI KLNRNLRTFTGVDLPS
TLIFDYNAAVLADHLRAELRGERPAVAA PVVRDVSDEPTAIVGMSRELPGGASPE
LWNLVAGRAISGFPVDRGMDLGLHHPDPAHPTGCTYTRITGGFLHDAAPDAGLFGI
SPEALAMDQQRLLSTSEALERAGVPLSARGSDVGVTGTFVHVDIYVTRUREVEPE
DVQGYTWTGASVGRVAYVFGFEGPVTDTACSSSLVAIHAAQALROGECSSWA
LAGATVWASPDALFESRQGLSADGCKRAYAEGADGTGWAEGVGVVVLERLSVARE
RGH"

ORIGIN
Query Match          9.4%; Score 47; DB 1; Length 3900;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 TGCAGGCGTTCCGCGACGCGCGACGCGCGCCGCTGTCGCGAGG 197
      |||||
Db 883 TGCAGGCGTTCCGCGACGCGCGACGCGCGCCGCTGTCGCGAGG 929

RESULT 5
LOCUS AB089954 58343 bp DNA linear BCT 15-FEB-2003
DEFINITION Micromonospora griseorubida gene cluster for the polyketide
            macrolide mycinamicin, partial and complete cds.
ACCESSION AB089954
VERSION AB089954.1 GI:28394148
KEYWORDS
SOURCE Micromonospora griseorubida
ORGANISM Micromonospora griseorubida
          Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
          Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE
AUTHORS Anzai, Y., Saito, N., Tanaka, M., Kinoshita, K., Koyama, Y. and Kato, F.
TITLE Organization of the biosynthetic gene cluster for the polyketide
JOURNAL macrolide mycinamicin in Micromonospora griseorubida
MEDLINE FEMS Microbiol. Lett. 218 (1), 135-141 (2003)
PUBMED 22472038
REFERENCE 12583909
          2 (bases 1 to 58343)
AUTHORS Anzai, Y., Saito, N. and Kato, F.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2002) Yojiro Anzai, Toho University, Faculty of
          Pharmaceutical Sciences; 2-2-1 Miyama, Funabashi, Chiba 274-8510,
          Japan (E-mail: yanzai@phar.toho-u.ac.jp, Tel:81-47-472-2072,
          Fax:81-47-472-2086)
FEATURES
     source          location/Qualifiers
     1..58343
     /organism="Micromonospora griseorubida"
     /mol_type="genomic DNA"
     /db_xref="taxon:28040"
     complement(1..910)
     /gene="mycF"
     complement(<1..910)
     /gene="mycF"
     /codon_start=1
     /transl_table=11
     /product="methyltransferase"
     /protein_id="BAC57022.1"
     /db_xref="GI:28394149"

     gene
     CDS
     5181..6443
     /gene="mycD"
     5181..6443
     /gene="mycD"
     /codon_start=1
     /transl_table=11
     /product="glycosyltransferase"
     /protein_id="BAC57027.1"
     /db_xref="GI:28394154"

     translation="MRIALTIGSGDQVQFPFALGVGLRARGHDVILGAPGEGIKHIVE
     RAGLRYRTPGDGFFTMEPVVDSRLRSPYLRLDLKALPAVEEDQDVDEIAAAG
     QGADAVVYAPLITAAAYGRSDVFAAANWNTRTWRFPAVSGQRRLGPLSLYHLF
     THARAAREWAMRPEVDGYRRQGLEPFGGRSPFLRLGRFPYLYPFSVLPKPRD
     WPARAHVYGVFWCDQWQPTRELVDLADGPAPVAADVROHLA VHREROVLVAYAAA
     RAAGRKI VVVDGDFVPPSDVLRLDHVDYRWLFPTTAAVHHGGGCTTAEVVRAGPH
     VVVPVFAHPFWARLTETGVAAPAVDFETLTREALVASTVAKVAREDSLGARAARLGE
```



```

/product="NanT2"
/protein_id="AAP42849.1"
/db_xref="GI:31044137"
/translators="MLGHRPSGLILVYVAQALRKKLAEKSTADRHPSGSAEM
ARLTFELPDGSTRVEDIVQVNLNAGYAGRSQDDVAHIAELAEGLVFTPTALYFVA
PYLAQIDRVAVGHRRTSGAEALVAGGDELLLTACDHTDRDLDEVGVAMSKNAG
PDVLAARWLADVEPLDILLTARWTRDGTETETIQHGTALLETLLPAYVWLSRSD
ALPTGTVLISGTPMTPTGVDQFADTWREVLGDEATGDTIRLAYDVHMPPEPIG"
3117..3818
/gene="NanT3"
3117..3818
/gene="NanT3"
/notes="putative two-component response regulator"
/codon_start=1
/transl_table=11
/product="NanT3"
/db_xref="GI:31044138"
/translators="MSTRCTGSLTNGWWSVELGIPLSVLVDHPPVVRAGISAWCAA
DAPISVVAEGANSVALHGPGRGADVVMDLLIQNGRPAYDELQELVAQERKVVYTM
RDSQDAALTCMDLGSATYITKABGQRHLVKAIIRAAEDIPYTPSLAGAFGSDTROSR
PVLVSREVELVWFQSEKXAVVAQSLGISERTVNTVLDVRVRIKYANAGPATTKAKL
VARAVODGLTALDEL"
3117..4923
/complement(3724..4923)
/gene="NanT4"
3117..4923
/complement(3724..4923)
/notes="hypothetical chemoreceptor"
/codon_start=1
/transl_table=11
/product="NanT4"
/protein_id="AAP42851.1"
/db_xref="GI:31044139"
/translators="MGEDFQLITRHOLGRAGGLVILVTLHIGSALPGLGLGLEY
RHWIPLAAYGLTLILGSGVLGLDGRPMPCGWPCALAGTFVAVATTSQVPADHY
FVSLSSAGSLAGFWVLLGHRGLVSGAFGLAHNAVTAALLITVDVPSRSIAGSNA
LSLSAGSCFQNTIAGALKLDSALGQALRAQERVTRIAVARQIQADQRRRIYEL
NATVPLLTGLAGSLDPEDEVRHACALRRLRFAESDCTFPLVHEMRACIDV
AENGTQVLAARGDPLEPLRLRALLDPVIAALAAAGTARVTVVRGGDQVRGVV
VDALRDLPEPKANGIRVTRVHAEDRLLEAAACRIATARPELTSRPEPPWPPSPS
AGRRSRT"
3117..6106
/complement(4916..6106)
/gene="NanT5"
3117..6106
/complement(4916..6106)
/notes="putative two-component sensor histidine kinase"
/codon_start=1
/transl_table=11
/product="NanT5"
/protein_id="AAP42852.1"
/db_xref="GI:31044140"
/translators="MSTISLSPGIDPPASARLPQSAGPWAREVYVFNRLALYGRAGV
MSACAVLGAAMPMEGAPAAVAAVLAWSLHLRLATTSAMRQRLPLALDITVMTA
LCLSQREPTPEACTHGTWILVSVFTAVAYGIMQPLLGLITLWLCDDVVGITAL
HPNLEWSGPEVTVWNTALAKAVLVVRSRDEADEAARARQQGGEAAEARC
AADEHLAHLADTACLTLTASAPWASMAETLRQAARDLLRLRAEQPVAVRVLAA
ELUDEAAHPLRVINFDVBLGTWPIAALRGGLGEALRNVARVAGVDVAVTAEER
VDHVIVTIADSGVGDPDEPHIPGHRIGIKRSIHGRMNWVGGRATVESRPGHGTVRL
WPRG"
3117..7493
/complement(6768..7493)
/gene="NanR1"
3117..7493
/complement(6768..7493)
/notes="putative regulatory protein"
/codon_start=1
/transl_table=11
/product="NanR1"
/protein_id="AAP42853.1"
/db_xref="GI:31044141"
/translators="MGLRVVDEKSGSLPRAQKLAALVILLIRSDQVMAEOLTREI
WGLPRPRDPAIYVCISRLKFLSRPGQGNPLITRPGVMLRRASCECDPFRQQLV
ADGRKVCYRGEPAQALCFQALALMRGPAFGVNGTVPVAGFLKRLDEMBEACMS
AEAHLEIGHRELHVLGRSLVAFVFFURETPIRQIMLALYRSQKSEALGIYESARRVL

```

```

DREILGPEGRKRLRELRLHTVLAST"
/complement(7573..8334)
/gene="NanR2"
3117..8334
/complement(7573..8334)
/notes="putative regulatory protein"
/codon_start=1
/transl_table=11
/product="NanR2"
/protein_id="AAP42854.1"
/db_xref="GI:31044142"
/translators="MRYEILGPLRVVDRSSVTSISARKIEVLLAALVRADEVSSSD
LVKEIWTGRAPROAALHVSQRLKFLRPNRTPSVIVTARSPYVLRQLGDELDLPH
VFOGLVROGRARAGRAGTHACAEYEGLEFWRGPVLELDRDGAIVNGFVCWBAQARL
ECVESFEAGLEIGHREFVSYLYRIEELPLNEAFYRQLMLALYRSGCRADALHVVQ
SAPAVLSSELGLEPSSAALKKHLHVALVDAPELSAA"
8919..17627
/gene="NanA1"
8919..17627
/notes="NanA1"
/notes="PKS"
/codon_start=1
/transl_table=11
/product="NanA1"
/protein_id="AAP42855.1"
/db_xref="GI:31044143"
/translators="MRFRYGVNAGSSPSHAQOATSPVAIVGLACRLPGADPPEAFWR
LIRAGENAVVPDRLPTPEGSPYFAGLLEHVDVDFDAGFGFGLSPREATAMPQORL
MLELAWEMEDAGLPKNLAERTTAVTGAIDDIYATLLHRRRPNDIAITRHTMAGLH
RGLIANRVSHLLGLRPSLTVDAASGLVAVHLCESLRGEADLALAGGVNLLAE
ESNRMAEAFQGLSPDCRVTDFARANGFVRGGGWLKLLAAAVADGDPVYCVTE
GSANNMGATDGLTRSDAODTVRQAWORAASPAELOVELHGTGPVGPPIEAA
ALGDALGDAGDCHRDRAVDAPLRVGSTMVGHLEAAAGIAGLLTALSIHRRLLP
SLNEFNPICPLAELGLAVQTAAGPWPORRRLTAGVSFGMGTCNCHVLAEPAP
AVLPDPTAPGDVSATIPATPEWVWSAASPKALTAQAALYEHLRHGRVHPVDIGH
ALATITPRRAVILGRDELDLSDALASSETQTSVIRGAGGRVAFPLFGQS
QREGMEIYAAVYFADALREYCAHLDPMLDITDPLDLMFAEAPPDGPPLNRTAY
TOPALFAIVALYLVTSWGTDPHLMGHSVGEITAAHVAGLSLPDACLVAARGEL
MOSITAPGMAWAQATAEAGQALAEVAGGVGLAANVAPSVISGDREAVATRAW
RAGRKATVLKYSHAFSLHDLGLDRLTVAAGLTPAAPALIPVNSLTGGAATEAQL
RSPDIWADHARQAVRFDAGVRHLCDAAGVTFLELGDPSALUTGWARSAAGADARP
VAVAVQRGRPEAQSAQAQAVHGVGVDAWAAAFAGHETRAPLFTYAFQDRHWP
DGLDERGRRPSTSPVSTDEPVVDASPADRAASPGELLALVTHAALVGHNSP
DGIDPALTFKQLGFDLSAATLSELSAATDELPAITLFDHPTNVAWLAARHSG
OPTAAATAATPSMAEDPVAVAVSCRYPGVESGEALNELVDEGVAVCEPQDRGW
DLAELGRAPDGGSGSATGCGFLYCGADFAEFGISPREALAMPQORLLELSE
LIERAGIPASLAGSATGVTVGATVDYGRLEHETAEELDHLLTGSTPSVSGRVAY
ALGEGPALTVDITACSSSLVAMHAAQALRQCECDLAGGVTVMATPGMFTSFQR
GLAPDGRCKPFAAADGTGNSGAGLVLRLSDARRNGHQLAVIRGSVNDQGSAN
GLSAPNGSQORVIRQALANARLEPADVDAVEAHGTGTLGDPTEAQAILLATYGGORT
DRLPLWLSIKSNIGHTQAAAGVAVIKMVALRHGRLPASLHIDAPSPHIDWSGTV
RLLSEVDPGTDTPGSDRPRRAVSVSGISGCTNAHLLLEQADPHPEPTTSGGVV
WLSKATDALRAQALAEVWTAGAPSPASPSPASPASPASPASPASPASPASPASP
VSGTDDELLSLGRAVADGLAPASAGAPGVWVPGQSGMGVGLLSDSSVF
AARMAACEALGELFVMSLTVLRGAPGAPESRVDVLPCLMAVMSLAAVWESYGV
TPTVAHSGOETAAACVAGLSLRDGRARVALRQALRALAGLMSLGSAGAE
RFLADIAGAAARVAVAVENGPSVTVSGDPTQVAAVVAACEAGHARTIDVYAGH
POVDRDLATIRIDLADLSFGASDAVYSVARGQTEELDADYFTNLRQVRFASA
IDALLAAGRVRVFEVSPHVLIPALRECEEEAAVATVTLRRDQGGPQVVARALGD
GFVAGLAVDMSRWFVGDREAGDEGHRPTVELPTYPQRRYMLADPHDREGRGTAG
VGTTPAGHALLLSAVELADGGLVLSRLPGDAWVGAHTVAGVQLVPGVAVLDWLLIA
ADAGGASLELLRAPLELSGSPSEPSAGVLAQVAVAGPDESRELRISRRPAD
AGAGCQWCHVAGSLAPGPPAPADTGATVPWPPAGAEALDPAGIYERAEPRGYGV
PALRGVTVLRDGLADVADLVPEERAGGEGGADGDTAGFGLHPLVLDALQPALI
APEDGTGGGAGGEAKLWLPFAWGSVRLWATGARAARVSLDGGGDDVDBRELR
EYSDPTGAPVLSVAVSVLRVTVRQVREASGAAGLFDALDTPVPOPSGADDAAG
CVAVLGEAPTEPGDCTYTDLPALLAALDAGAPLPSVWMPRPAADPAGAPEDAA
LSARGVGAALRAWVAEPRLTVSRLAVTVRGAAGGAGEPVDLAAAAGCARGVQ
ABHPDRILVLDVDDVDMGADTDDIIGAAGLAAALGEQVVALRGTLAPLARSA
TPGGVAFDNGTVLTVDSGPLAGSVAEHLVRAEGVRHLLVRFEGADGQVITDRQD
AQVHWTVDFRDTAALREVAQVDPFAHPLTVVHVHAGLSADITSGAARGWAVAGVY
RALHQATLALPSVRFVTLSDATANDGPAPAPPEAAAGAFCAVTVDRRAGLGLDVA
FGFWAADDGDDGSGRWNTGVLGADRGLALLRAACRADRPRLVAADIRTRALTAPFA

```

HELPALRRTLIGASASAGGRAPVRRVAAAAPGRTTDWASRLVGLGPAERRAVLELV

Query Match 8.2%; Score 41; DB 1; Length 132544;  
Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 GCGACCCGATCGAGCGCGAGCGCTGCTGGCCACTACGG 464

Db 37467 GCGACCCGATCGAGCGCGAGCGCTGCTGGCCACTACGG 37507

## RESULT 7

SNAL132222 20394 bp DNA linear BCT 24-APR-1999  
Streptomyces natalensis pims1 Gene.  
AJ132222  
AJ132222.1 GI:4678702  
pimaricin biosynthesis; pims1 gene; polyketide synthase.  
Streptomyces natalensis  
Streptomyces natalensis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomycetes.

## REFERENCE

1  
Aparicio, J.F., Colina, A.J., Ceballos, E. and Martin, J.F.  
The biosynthetic gene cluster for the 26-membered ring polyene  
macrolide pimarin. A new polyketide synthase organization encoded  
by two subclusters separated by functionalization genes  
J. Biol. Chem. 274 (15), 10133-10139 (1999)  
99214571  
10187796  
2 (bases 1 to 20394)  
Aparicio, J.F.  
Direct Submission  
Submitted (12-JAN-1999) Aparicio J.F., Microbiology, INBIOTEC, Av.  
Real, 1-Leon, 24006, SPAIN

## FEATURES

1..20394  
/organism="Streptomyces natalensis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:68242"  
1..20394  
/genes="pims1"  
1..20394  
/genes="pims1"  
/function="pimaricin biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase"  
/protein\_id="CAB41041.1"  
/db\_xref="GI:4678703"  
/db\_xref="GDB:Q9X993"  
/db\_xref="SPRENBL:Q9X993"  
/translation="MSNEKRLKRLKRAJADLHETRQQLDEBAKREPLAIVSMACR  
FPGVRSPELMDLGDVAVSSFPNNGWDLALYHSDPAHQGTSYAREGEFLDHA  
GDFDCEFGISPREALAMPDQRLLETAWEAVERAGIDPESLAGSRTGVFGTGHG  
YDAEGRRADEVGHLTGNIHSIAGRSYVLGLGSPALTDTACSSSLVALHLMH  
LURSDAVRNGYFVLAVSANGQDGAENGLAPGRCXPPFAAADGTGMSGGLL  
IDAVEAHTGTTGLDVEAHLATYQQRAADRPCGLGSMKSNIGHTQAAGIAGV  
KWLVAHGHLPRTLHDEPTGHVDWSEGNRLAEPWPSPGRPRRAVSSFGISG  
TNAHVILEQAPAEAPPAARPGALPWLISARTAGLRAQADRLGRLHRLDRADLE  
PAVAHALADTRILMEHRAVVAAGDEEPLRLDAAAGTANGLVGVAKVAASFL  
FAGCGORPMGRHLLHAHVAFATAFADVACEDLPHLDRLPRLDIIVAEESAAALID  
QTYATQALFALTALFLVSWGVAIFRFVAGHSIGELTAHVSGLVTLHDAARLVAA  
RTYLMQALPAGANVAQVATEIRLEAGHEDHVALAANGPDSTVSGDEQATVEI  
AAHWEAGRTKRLVSHAFSPHMDLDFRRVARGLTTHAPRIVPVSTVTCALAT  
EDLRSPDWVROVRETVCAAVRTLEAGVTTFVEIGTGVLTVMQDCLTLEP  
VLPLRTGRPTVALTEGVATFVGVFVDRSAFPGACPTSRADLPITYAFQRMWYL  
DPADHDEGEAAAAGAGFAWVEREDLQELSAVLADGSEADSLGSLFPLTSSWR  
HTDRELPADEHPDAMGRHPVDVGLCILDALDERPSCPPYRRLAATNAARPEGA  
GIAQLVLCVTRGAVAVDRHALKSPLOAQTWGLGRVALESQSGGLIIDLPNIDGR  
AVSALLSTLAGEEDQAVRPAVGFARLERITPGGDGDRWSTHGTVLVTGGTALCA



APDAAEDEADGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRTIDVAYSLATGR  
SAFHRAVILANRADLHLSALAEHGTEPAVVAODRARGKLAFLFSCQSORLG  
MGRELXGYPFAFALDAVCAALDAHLDVLDVWGDALLENARTGVAOGLFAIEV  
ALFRLLSWGVRPHLLHGSGETAAAHVAGVLSIPDCAALVAARGRLMQOLPSGGM  
MAIRATEDVPLHAEGLSAAVNPSSVVSAGAEDEVIALAAHFEERGRKTRURVS  
HAFSPMLPEMLADFRVAVDGMTAAAPRIPVSVNVTGRPATAEELCCAYWVGHREA  
VRFADGVALREQVWTTLELPGDGLSALAAESAAADSVLAPLRKNRPAPALLTA  
LHLLHAQGTVDWGAAGFAGTARWDLPTVAFQHERFWPSCGAGARAGDVRAGLSGAG  
HPLLAGVELAGSGRLTGRSLSSHPMLADHVVLGSLVPGTALMELVLRADDEVD  
CAAVDELTLAIPVLIPASGAALQVQVWVGEDEAGRRPVSHAREGEGPTHADGAL  
APAEPTFDTAIWPQGAHDDAAGCYERFADAGFAYPVPTQGLRAAWKLGEDIYAE  
VALPEGTGNGHLPALFDAALHAALGSGGTDEAAVFPFWNGVTLHTAGASVRVR  
IRPTEGTSIALVDTPAGVAVSLTARPITAGLOQTDGDRSLFQVDMWTLHTDER  
ANSIALCKGTGEGIDTSLQPHADLDLAATGVHDVTLAIPLTRTACTGVESVRAATT  
GALATRSMLADDDPAAARIVTRTGA VSGTDLACASVWGLVRSALLHPHGRFGLVDV  
DYDQDAEVLVPRALASDEPQVLRGGEVLAARVRAQSSDTWDPGTVLITGCTG  
GLGSRVSEHGVRSLLVSRGPAAGVGDVLAELAECAQVTEACDVTDVA  
VADLVARRHISAVVHTAGVLDGCVVESITPERLSAVLRPKVDAAWNLHEATRGDLDA  
FVVFSSVAGTFSAGQANYAGNFDLALAYHRAVGLPVPVSLAWGPMQDGMGTGL  
SDADVOTAROCMPLTVEEGIALFDAALGSAEPMALPVRLDLAALRAQCPOLLRIG  
LIRTPGRTAAAETGDTAAAFAGRLTGLSRAEGREVLVRSOIAQVGLHABATEI  
QDRAFLDLGDSLTAVELNRLGAVTGIRLPATLLDYPPTPAELVAHLHARIAPEPT  
VGPRLAGELDERMEKSGGLDTEENHEQIAGRLVLRKAKMDALRDTAAAGHDGSPS  
DEDFEFASDDDEVFDLLDNEILGLS"

## ORIGIN

Query Match 7.2%; Score 36; DB 1; Length 20394;  
Best Local Similarity 100.0%; Pred. No. 3e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 121 CGGCGAGCGGCGCTGGCCCCGAGCGCGCTGCAAG 156  
Db 5320 CGGCGAGCGGCGCTGGCCCCGAGCGCGCTGCAAG 5355

RESULT 8  
AX067996 20394 bp DNA linear PAT 19-JAN-2001  
LOCUS Sequence 3 from Patent WO0077222.  
DEFINITION AX067996  
ACCESSION AX067996  
VERSION AX067996.1 GI:12329806  
KEYWORDS Streptomyces natalensis  
SOURCE Streptomyces natalensis  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Martin, J.F., Aparicio, J.F. and Colina, A.J.  
TITLES Genes encoding enzymes in the biosynthesis of pimarinin and the  
application thereof  
JOURNAL Patent: WO 0077222-A 3 21-DEC-2000;  
DSM N.V. (NL)

FEATURES Location/Qualifiers  
1..20394  
/organism="Streptomyces natalensis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:68242"  
1..20394  
/notes="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC24642.1"  
/db\_xref="GI:12329807"

## CDS

/translation="MNEKRLRYLKRATADLHETRQQLDETEAKOREPLAIVSMACR  
FPGVRSPELWELLRDGDVASSFFPNRNGDLDALYSDPDAHQGTSYAREGFLHDA  
FGEDPFGFISPEALAMPQORLLLETAWEAVERAGIDIPESAGTSGTGVGTGHHG  
YAEGRRADEVCGHLLTNHISIASGRISYVLGLEGPALTVDITACSSILVHLAMH  
ALRRDCAMALVCGATWMTSPQVRSRQGLAPDGRCKPFAAADGTCWSGVGLL  
LVERLSDAVRNGYPVLAVLKGSAVNODGSNGLTAPNGSQQVIRICALTGGLAASD  
IDAVEAGTGTTIGDPVEAHALLATYQQRAADRCPGLGSKSNIGHTQAAAGIAGM  
KMLAVRHGHLPTLHDEPTGHVDWSEGNRLLAEPFMPWSPRRRAAVSFGISG  
TNAHVILEQAHEAPAPAPARPALPWLISARTAEAGLRAQADRLGRHLRDRADLE

WNLHETRGDLTDLAFVFFSFGSQAQYAGNAFLDALAYHRAVGLPAVSLA  
WGPWQDGMGTLSDADVQIARQGMPLTVEEGLALFADAGSABPMAVPLVRLDLA  
ALRAQGEQPIRLGLRTRRSRGAASAGIAQRLAGLSTAEERREALLDVVRAQIATV  
LGHAFETIADPRAFOGLDGLSLTAEIENLNLGKATGRLRATVDFDPTVDALAAHL  
LDELFGAETATGATPPLVPLGLSLADPFSVIVGMSRFPFGVASPDELRLWRLADGVD  
AVSAPFTDGRWEIDTIDYPEREGATATSGGLFHDHDAEPDFPFMGSPREALTTDAQQ  
RILLLETWEALRAGMDPATLRSRTGVFAGVMDHDTLLSGRFEFGYQSGSAGSV  
ASGRVSTFGPEGPATDTCSSSLVALHAAQSRGSECLALAGGVVMTSTPLTF  
VEFSRQGLSADGRCCKAFADADGVMAEGAGILVLERLSADRNGHRLATVRCGSA  
NQDGSAGNLTPNGQAQRVIRIQALASAGLSADVDVAEAGTGTGLDPTIAQALLA  
TYGQRPEDRLPILSVKSNIGHQAAGSAGVIMKVLMMRHGVLPRTLVHDEPSSHV  
DMSGAVELLTSAEWPEGGERAGVSFGISGTNAHVLEPQPEVAETESITPTD  
APDAEADSGSTGTPALLSGSALRAQAARLLSLRDGDPGRITDVAYSLATRG  
SAFPHRAVTLAANRDLHLHSIALAEHTEAPVAVQDRKARSGKLAFIPSGOGSRLG  
MGRLEYGRYPAEALDCAALDAHLORPLRDVLTWGEDAELLNRTGVAQTCGLPAIEV  
ALFRLLSEWVRPHLLHSGISEIAAAHVAHVLSLPDACAALVAAGRLMQOLPSSGM  
MAIRATEDEVPLHAEAGVSLAAVNGSSVVVSGADEVIALAAHFEERGRKTRKVS  
VAFADGVGLMPLADFRVADGMTYAPRIPVVSNVTGRPATAEELCCABYVWGHREA  
HAFSPGLMEALREQGVTTIFELGPGDSLALAAESADDSVLAPVLRKNRPEAPALLTA  
LARLHAQGTTPVDSAAFAGTGRWVDLPTAFQHERFWPSGGAARAGVRSAGLSGAG  
HPLLAGAVELAGSGRLITGRSLSSHPWLAHDVILSVLPVGTALMEILVLRADBDV  
CAAVDELILAEVLVLPASGAALQVQWVGEPEAGRPVSVHARSGECPWTLHADGAL  
APAAETPDDTAIWPPQGAHEDAACYERFADAGFAYGFPVQGLRAAKWKGEDYAE  
VALPCTGDNAYGLHPALFDAAHAAALIGEGEDTEAEPFSWNGVTLHATGASRVVR  
IRPTGEGTIALVDTAGAPVASVRSILRTAPITAGQITQDGRDLSLFTQVMTTILHLDER  
ANSLALGKDTGEGILDTLSIOPHADLDLAAATGVHDVTLAPLPTTAGTVEGVAHAAT  
GALALIRWLAADRFAASRLVFTVRGASGTDLGASVWGLVRSALHLEHPSGLVLDV  
VDQAEVPLVRALASBPQVIVRGGEVILARLVAQSDVTVWDPSPGTTLTGSTG  
GLGRSVARHVSSEHVSRLVSRGPAAGVDALVAELAECAQVTVACBDVDAVA  
VADVARHRSISAVTAGVLDGVSLETPERLSAVLRPKVDAANLHETRGDLDDA  
FVVSVAHRTGTSAGAGNAAGNAFLDALAYHRAVGLPAVSLANGPNSQDGMGTGL  
SDADVQIARQGMPLTVEEGLALFADAGSABPMAVPLVRLDLAALRAQGEQPIRLG  
LIRTPGRFAAATAGTGTAAAGRLTGLSAAEGRVILGAVRSQIAGVYLGHAEATEI  
DQRAFIDGLFSLFRAVELNRILGATVGRILPATLLFDPTTAEILVAHLHARLEPPT  
VDEPFLGLEERKEKSFGLDITEENHEQIAGRLVIRAKWDRALDRTAAAGHDGSPS  
DGFDFESASDDVEFDLLDNLGLS"

ORIGIN

Query Match 7.2%; Score 36; DB 6; Length 20394;  
Best Local Similarity 100.0%; Pred. No. 3e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 CGCAGCGCGGGCTGCGCCCGAGCGCGCTGCAAG 156  
|||||  
DB 5320 CGCAGCGCGGGCTGCGCCCGAGCGCGCTGCAAG 5355  
|||||

RESULT 9  
SFU78289  
LOCUS  
DEFINITION Streptomyces fradiae tyactone synthase, starter module and modules  
1-7, (tylg) gene, complete cds.  
U78289  
U78289.1 GI:2317859

Streptomyces fradiae  
Streptomyces fradiae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 43280)  
DeHoff,B.S., Sutton,K.L. and Rostock,P.R. Jr.  
Sequence of Streptomyces fradiae tyactone synthase gene tylg  
Unpublished (1996)  
2 (bases 1 to 43280)  
DeHoff,B.S., Sutton,K.L. and Rostock,P.R. Jr.  
Direct Submission  
Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate  
Center, Indianapolis, IN 46285, USA  
tylg.

COMMENT Location/Qualifiers  
FEATURES  
source  
1. .43280  
/organism="Streptomyces fradiae"  
/mol\_type="genomic DNA"

/db\_xref="taxon:1906"  
1. .43280  
/gene="tylg"  
816. .14234  
/gene="tylg"  
/note="ORF1; putative"  
/codon\_start=1  
/product="tylactone synthase starter module and modules 1  
& 2"  
/protein\_id="AAB66504.1"  
/db\_xref="GI:2317860"  
/translation="MSSALLRAVQSNCGYDMLTNTAAQNTQDQDDVDVDPSTHGGE  
TAVGMSCLFGAAGVEFWELLRSRGMPTQDDGTWRAALDEHAGFDAGFAGMNA  
QAAATDHRMLMELGWEALDAGIVPGLTGTCTGVFAGVADSDYAVLTRSASVAG  
GYTATGLHRLAANRLSHFLGRLPSLVDSQASLAVQLACSLRGETSLVAG  
GVNLLITESTTVMERMGALSPDGRCHTFDANGVRGEGGAVLKPDLAALADGD  
RYVCKGAVNNGGGASLTTPDREAQVLRQYRRAAGVSTGAVRVVLEHGTGTRA  
GDPVEAALGAVLGADSGRSTPLAVGSVKNVGHLEGAAGIVGLIKATLCVRKGE  
VPSLNFSTPNPDIPLDLRLVYTEROEWEEDDRPVAGSVSGFGMGTNVLHVI  
PAAAGSGAGSGAGSAGISAVSGVVPVVSRSRVVREAGRLAEVVEAGVGLA  
DVATMADRSGFYRAVLARGEALAGRLALAGDGDAGVTVGAVLDGCVVVGAA  
CGAGAGGAGGAGGAGGAVLVFPQGTQWVMGAGLLGSSEVFAASRRECVLAV  
HYWDLLEVVSGCAGLVRDVPQVTVAVMSLARYWQMGVDAVAVCHSGEIAAA  
TVAGALSDEAAVVALRAGLIGRYLAGGMAAVPLPAGVEAGLAKWPGVEVA  
GPATVSGDRRAVAGYVAVCAQEGVQARLIIPDYASRSHVEDLKEGLERFTEV  
VLRVHAETAEADRSVHATGLRRQDPSRHLTSTAAMAHGATLWDPALPGHL  
TTLTYFNNHHWLDITDGGGDDATOEKSGPLTRELRGLPSSOKOGLFLDLVCR  
HTAVLGLTAAEDVDLDFKQCGQISMTQKVBELNRLLTETGLATLTVYDPTPRA  
LAQVILTELLDGSFSGVLAQPAQFAEQPIAVVGMGRFPGVSGSPALMRLVVEG  
VDVAPFPDGRMGVEGLYDPPEVAGKSYVREGGLHDAEFAEFFGIGSPREAVM  
DPQORLLTWEATIERAGIDPHSLHSGRTGVAGVMPQVGYFRLAEGAGSDGLLT  
GTSFVSGRAVTLGLEPAVTVDTACSSILVALHLAQLRGGCDMALAGGVTVM  
AGPQVFSRQGLAAGRCYAFADAGTAAEGAGVLRVLSRDLRGLHPLVAV  
VCGSVNQDGNGLTAPSGSQERVIRALGNARLTVADVDVVEAHGTGRLDPIE  
AQAALITGDRDGGRPVWGLSGNIGHAAGAVGVIKMWLRYGWLKPLRLHVD  
EPSRHVDMSAGVWLLTEARFPGVDRPRAAFVSGTNAHLITAPDTEAESAA  
TTPVRSVSESAVLDARSGVVPVVSRSRVVREAGRLAEVVEAGVGLADAVT  
MAGRSFGRVAVLARGEALAGRLALAGDGDAGVTVGAVDPTGSGGGVLFV  
PQGTQWVMGAGLLGSSEVFAASRRECVLHVEWDLLEVVSGAGLVRDVPQ  
VTVAVSLARYWQMGVDAVAVCHSGEIAAAVAGALSDEAAVVALRAGLIGR  
YLAGMAAVPLPAGVEAGLAKWPGVEAVAVNGPASTVVSGRRAAGVAVCAQ  
GVQARLIIPDYASRSHVEDLKEGLERFTEVSAHPVLVHAIEQTAEDRSVHATGL  
YVNRNRNRFSAVVGGLLEHRRFIEVSAHPVLVHAIEQTAEDRSVHATGLR  
RQDPSRHLTSTAAMAHGATLWDPALPGHLTTPYFNNHHWLDITTPPTAT  
TQSPDANRYRVTWALTESSPVRPHSICRLLVAPPTDGLLGLLTLVLSERAS  
VARLEPITGARRAEVLELKPSSMESAGENTTVSLLGLVLPSTDAVTSIALQVSD  
IGVPAARVVALTRAVAVPGETPQDAGAQWGRVRAVLELDPIDWGLLIDLPETAE  
TRTPETSPQPTPERLPQTPNRALALAAVLAGRDGQVAVRAGIYGRVSRRAA  
AGAAWQSPGTVLTGGMGAIGRRLARLAAEGERLVTSSRGPAPGAALAEELR  
GHCEVTHAACVAERDALAALVAVPNAVFTAGLDDAVDTLSPESFETVRGAK  
VCCAEILLHQLTADIKGLDAPVLFSSVTGTWGNAGQVAAANAALDALAEERRAAGLP  
ATSVAMGLWGGGMAAGAEESLRRLGRAMDPAVDALLGAMGRDVCVTVVDWM  
ERFAPATNIRPGLFDTPVPEAREALTAGTTPDGAPELLARLSMLNETBRLEKL  
VELVTEAAFLVRLPNTDAIGERPFKSAGFDSLTSLERLNRNLGTLKLPATVIFD  
HPSPTALRLLDLRTAGAPADAPPLPVAADDPPVIVGMACRFPFGAGTPEA  
LWKLVTEDRVIGAAPTDRGMDLSDVVDPEPGVAGTYVREGGLHDAEFAEFFGIG  
SPREAVAMPDQORLLTWEATIERAGIDPHSLHSGRTGVVGLTHQEYASRLHAPE  
EYEGYLLTGKSAVSRSYITIGLEGPUSLIDTACSSLSVLAHNAQAALRGCECDMA  
LAGCVTVMAAPGLFVFSRQGLAAGRCYAFADAGTAAEGAGVLRVLSRDLRGLH  
LGHPLVAVCGSVNQDGNGLTAPSGSQERVIRALGNARLTVADVDVVEAHGTG  
TRIGDPLEAQALIGTGRDRAFCPVLGSLKSNIGHAAGAVGVIKMWLRYGWLKPL  
LPTLHVDPSRHVDMSAGVRLTAREWPGVDRPRAAVSAFVSGTNAHLITLLEAP  
EALAEALDAPAEAPERAPDVTDTLEAEADTAEGAKAPGSPEREAQVAVVP  
VVVSRSRVVREAGRLAEVVEAGVGLADAVTVMWAGVTRGTRAVLARGELAG  
RLRLAGAGDDAGVTVGAVDPTGSGGGVLPVFPQGTQWVMGAGLLGSSEVFAA  
SMECARALSHVHEWDLLEVVSGAGLVRDVPQVTVAVMSLARYWQMGVDAVAV  
VGHSGEIAAAVAGALSDEAAVVALRAGLIGRYLAGGMAAVPLPAGVEAGL  
KWPQVQVAVNGPASTVVSGRRAAGVAVCAQEGVQARLIIPDYASRSHVEDLKE  
ELERLVSIGTRPSRFPVPCSTVAGEQPGVAGVQVQARLIIPDYASRSHVEDLKE  
HRRFIEVSAHPVLVHAIEQTAEDRSVHATGLRQDPSRHLTSTAAMAHGATL  
TWDPALPGHLTTPYFNNHHWAVTSPAGVDAAGAAFGMTWEDHPFLRGLPLA

```

DSGRVFAGRLAGSHDLTDHVAHSVGLTLPGLTAFVBFALHAGAAATCGRLEELSVEA
PLVLPAAGGVQMRVSAADESGRRVAIHSAPAAVHSAEAGDSAGVWTRGEGTLL
VPDEPTPPADADWAPRAGERVEPAELVERFGALGVGEAGPAGVAVWQOPDALL
AEVLLPRASTAGRGVGHFALLDALQPMWAGLLSVBDEAVLLPFAKQGVLSIYATG
AGALVRLTKAGDGVSLQADTSGAAVLSLGAIVMRPLAKRKLDVLLGIDAGERSLY
RVENQRIIPAGPPRSWAVLGPDAIRLGTGLGDQPDGGPTALYPEVRALRKALAAV
APREAVVVLVLSGAGATPESVOTTERCLTALQDMLDABELVDTPLVITRGAAGAV
PGEIIGLACAGVWGLVRSRHPGFALVDTGHPDDRTALPLARVLIDGAGQOLS
LRAGTAPVLLRAGTPEEROGPAFDPAQTVLNTGCTGLRLLARHLAAEHGVHLL
LLSGRAAGADLAELAGLEPCFPAACDAADREALARYLAEVADRPRTGTVIHA
AGVDDTGLDALPRTGIVWVRKADALNLHLEITRSPUSVFAVFGAAGLGRPGQ
ANYAANTFDLALAHRAEDPLLVPMRLDLAALVREAEHGDPAVFGPGLLGLPARAAV
THDQCLAFRFGAPAGGEGTAERLAGLGEARLRELRLVRAEYGVLYSGPDAVEP
ROAAVLDGAPALAEGLPPLVPMRLDLAALVREAEHGDPAVFGPGLLGLPARAAV
GRPFKIDGFSLTAEVRLNRLGAATGLRLPTALVFDRTPTQAVAEYLAAELAGPRDGG
DTAAAAFEGLEBALAAVGAALDRLRLRLRLTELAALTPQGRNPSAPAPASDL
DERLDSANDDLFAFIEQL"
/gene="tylg"
/function="encodes starter module"
942..2156
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase KS-Q of
starter module"
2571..3557
/gene="tylg"
/function="encodes acyltransferase domain of starter
module"
3675..3929
/gene="tylg"
/function="encodes acyl carrier protein domain of starter
module"
3993..8471
/gene="tylg"
/function="encodes module 1"
3993..5264
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase of module 1"
5631..6617
/gene="tylg"
/function="encodes acyltransferase domain of module 1"
7410..7949
/gene="tylg"
/function="encodes beta-ketoreductase domain of module 1"
9220..8471
/gene="tylg"
/function="encodes acyl carrier protein domain of module
1"
8541..13970
/gene="tylg"
/function="encodes module 2"
9541..9812
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase of module 2"
10260..11246
/gene="tylg"
/function="encodes acyltransferase domain of module 2"
11319..11876
/gene="tylg"
/function="encodes dehydratase domain of module 2"
12861..13415
/gene="tylg"
/function="encodes beta-ketoreductase domain of module 2"
13719..13970
/gene="tylg"
/function="encodes acyl carrier protein domain of module
2"
14351..19945
/gene="tylg"
/notes="ORF2"
/codon_start=1
/transl_table=11
/product="tylactone synthase module 3"

```

```

/protein_id="AAB66505.1"
/db_xref="GI:2317861"
/translation="MTAEVLDAVTRKRLGALEREAREPTAVVAMSCRYGVTTPEDLMR
LLADERDAVSLDRDMDALYDPDGGPQTSVAREGGFLSHCGDAEAFGLSPRE
ALAMDPOORLLTSTSEALERRAGVTADRAAGSTGVYAGVMDYDYGARVLYGAGPP
EDLEGYLVNAGSAGISGRVSYTFLGRLPQAVTNTACSSLSVLSHLAVRANGECOM
REGRPLVAVIRGSAVNQDAGNGLTAFNGRAQKVIEDALRDAGVQDQVDAVAHGT
GTBLGDPIEAGALLATYGTARTAEPLWGLSKSNIQHTQAAAGVAVIKMVLAMRHT
RLPRTLVHDDPTVDWEKGVRLLTPEVPVPEAGEPRRAGTSSFGASGTNAHVLE
SVFAGEPPAAGRPEDTGGANTVSRGPAALRAQARLYDALTGTCGTGGCGAGGAGP
GTAEVAGALAHARTAFRRRAVVLGCONRAELLAGLRELAEHEHPQPRVVTGTAPATER
TAFILFGQSGRAGSGHLYRRHPVFAALDEVCALPEHLRHLRDLMLFAEPSPER
EPIDRTFTQFALQATLFRLEAHHGIRAEALCGHSVGEIAAAHAAGVLTLPDAAR
LVAAGRLMALPAGGAMAALRAEIEIAPLERRAGELAAVAANGSSVSVVSGDEAA
VLELLEOWRAEGREARLVASHAFSPHSDMGMLTQFORVARTLTFAPTTLPVLTLLTG
TPVTEETLCTADHWROAREPVLDMWTLRADGIDTFVELGPDGVLMSAMARDCAAD
RPGDITTGAGDGETPDPFLLTLPRLRSVPFTGDAEHGGFERALATAYAHGVPLRLAP
APDAASLAAVELPTAFQTHWYMDAPAAPALPAGDDAGHPLLSAALDLPQGRGT
VMTGALSAATLPWAADSHVGRVTLPGTALLDLALHAAPRGELTFEAPLVPEDGEV
RLRVLAEPDASGVRELSVHSGEDGGTWRHATAVLTDTTGTAGEPAGAPAAWPPGD
AEPDLAAEYERFADAGIGYGPAPRGLRSARWDGDAIILADVRLPGLAGADRYCIHP
ALIDALHTAAALGAGHGMPLPTWNGVTLHARGAHAIIRVRLTPAGDPAVAVTVDPA
LRGETPHHPLASALADTGPAPRVLADLKGIGTATPRELASQAMALVQQWADAEEA
EGRLVTVRRADVIGDVTDPAAATVWGLVRAAQSEHPGCFALLDDTSPRSQULLPR
VAGTAEQLALRDGTLTLPSTLRTATPAGARLPALDGTVLITGTGSLGAEARHLVTR
HGARELLTSSRGQAPGAELVAELAAAGHADVACDVADRALGALLRALLRVPAGHP
LTAVLHTAGVLLDGVLTQPTORLAAVLRPKADAVRNHLELTQGHALSFAFLYSSAAG
VLGSAQSGVYAAANAYLDSFAVWRSLRGLPAVSLGWGPDGGSAGSLGTTARLR

```

Query Match 7.2%; Score 36; DB 1; Length 43280;  
Best Local Similarity 100.0%; Pred. No. 2.8e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCCGAGCGCGCTGCAAGCGCTTCGCCGAC 168  
|||||  
Db 15074 CTGGCCCCCGAGCGCGCTGCAAGCGCTTCGCCGAC 15109

RESULT 10  
SNA278573/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS

SNA278573 84985 bp DNA linear BCT 07-FEB-2003  
Streptomyces natalensis pimarcin biosynthetic gene cluster.  
AJ278573  
AJ278573.1 GI:12055067  
ABC transporter; cholesterol oxidase; cytochrome P450 monooxygenase;  
ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine  
dehydratase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE  
gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK  
gene; pimL gene; pimM gene; pimN gene; pimO gene; pimP gene; pimQ  
gene; pimR gene; pimS gene; pimT gene; pimU gene; pimV gene; pimW  
gene; pimX gene; pimY gene; pimZ gene; pim1 gene; pim2 gene; pim3  
gene; pim4 gene; pim5 gene; pim6 gene; pim7 gene; pim8 gene; pim9  
gene; pim10 gene; pim11 gene; pim12 gene; pim13 gene; pim14 gene;  
Streptomyces natalensis  
Streptomyces natalensis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomyceinae; Streptomycetaceae; Streptomyces.

REFERENCE  
AUTHORS  
TITLE

Aparicio, J.F., Fouces, R., Mendes, M.V., Olivera, N. and Martin, J.F.  
A complex multienzyme system encoded by five polyketide synthase  
genes is involved in the biosynthesis of the 26-membered polyene  
macrolide pimarcin in Streptomyces natalensis  
Chem. Biol. 7 (11), 895-905 (2000)

JOURNAL  
MEDLINE  
PUBMED

20547809  
11094342  
2 (bases 1 to 84985)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Aparicio, J.F.  
Direct Submission  
Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics,  
Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,  
SPAIN

FEATURES  
source

1..84985  
/organism="Streptomyces natalensis"



```

/mol_type="genomic DNA"
/db_xref="taxon:68242"
<!-- 2086
/notes="ORFX"
/codon_start=2
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC20917.1"
/db_xref="GI:12055068"
/db_xref="SPTREMBL:Q9EWA5"
/translation="TRPDAREATEPSPYGALLAPERSADRVETGKWTSDWSQARFL
ADHAYCEVLLTAGHPVLLLEQWADTASDLLRLSPRCQGHPLGIVLTARKEE
TESDATTREIVSDVGRPTATRLCGLPQAVGLVEAHAGSGVDPEVVEVLRHSE
GNPYFMOLISVQDARRLRPDADALLAHIPTGVREALNORFAALPEPVLVRLCL
AVGAEDTDLRRRTATCDASVAGLESARAGLGEDPHHPQLHFALVQELTFD
ELPADRRLLKATADALYARRGHVGDSEIERVAHSHWHAKSLPAARVLPLLAL
EHAQMAQEYKEIMURRAHLVGLFPDDPATVRLLEQKHLQLGQVLATRGYGHPE
AETALTRGALSTVTHAPEDPSPVLMALCAASLVTRYDASROLGRLRLSWQTRFPV
AVLGAAYGEGIVLHVRGMLPQALAELEHGVREADHFANRGHALARTFQDHPRVCSRY
NPTFWLLGHQIADARRRELLSLTEYGRSPDRSPALYVDGVVAWEGQRTYLRHPL
HIGLQQAQHNGQIEEARTTFRMLLAAVEQRRERYVLHPALPATLILHELLHGHGAE
ALADRDPRGRQV"
2190..3566
/gene="pimK"
CDS
2190..3566
/gene="pimK"
/function="putative glucosyl transferase"
/codon_start=1
/transl_table=11
/product="pimK protein"
/protein_id="CAC20918.1"
/db_xref="GI:12055069"
/db_xref="GOA:Q9EWA4"
/db_xref="SPTREMBL:Q9EWA4"
/translation="MESARPIILFVSLPESGLNPLILVLAGELSRQGVEDLWFATDEP
RNDVRIAEQSPVEFASIGVDEMSAVTMSDEVREYTPQSRFAHRAVVRHTYRP
GLAEFRRLQAVIDVQALMVIDICISGFVADAAIARNIPYVLSVPLFNSVLTAT
HPAKSVTPRGFPVPTGLSRMTLAQRVANELLFKLRTFAMFLNPLKGLNDDNRN
ELGLXPASPMARLEHADVLNLSAELDPDPIPEKMLVGMVPLPEAPDDQDL
WLDQSSVTVGLGTITRLTREQVGSMEVARELDRHQVLWKLPEQQLHLIPRESL
PGLNRVSEWVPMQMDLAHPVFFTHGGGNGFNEMYPFGPLVLRPLWMDYCDQAV
RGQDFGLSLTLDSPQITIDVNDVVDKLTVLGTSPFYEKARRAALRSAGRETAAGL
VLSLPLA"
/gene="pimS4"
complement(3662..9736)
/gene="pimS4"
complement(3662..9736)
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="pimS4 protein"
/protein_id="CAC20919.1"
/db_xref="GI:12055070"
/db_xref="GOA:Q9EWA3"
/db_xref="SPTREMBL:Q9EWA3"
/translation="MENDKNDKLVLDKXVNTDHLQTRLRREVESAKOEPATIVG
MSCRLPGCVSRSPFEWQLLXDRGDGISTFTDGRGWDLDLAGGSGSSATAEGGDTA
ASFDAGFGISPREAVTMDPQOILLLESSWEALERAGIDVLSRGITSTGVFGSGID
YINTVMSREDIDEGHSTGLTASILSGRVSTYFLEGPVAVTDTACSSLSVLHLAQ
ALRVGECSLAGAGVIAISTPMSFVGF SRQGLAPDGRCKAFSDTADGTAAWEGVYL
VWRLSIDARNHGHVLA VRGSAVNQDAGNGLTAPSGPSQOQVIRQALAAAGLSAD
VDVVEAGTGTALGDPEAOALLATCYDRDPERPLILGSKNLIGHTQOAAAGAAGLI
KMTLAMRHGVLPRTHLVEARSSVHWSGALOLLTEPQEPKADRPWRAGVSSPAISG
TNAHILIEQPAAEPEDEGPHTPSPAPVLPVLSAKSEERALTQGLRISALTRSPID
LGLSLATTRALLERAVLLEGGTVEVARGHAETAGKLAFLVPSGGQAGVRNGREL
YGRFPFAEALDGLAHDFADGLRDVMEGEAEGDDETGTQFALFAIVALFRLESIG
VRDPFVAGHSITAAAHVAVGVFLADACALVAARLMQALPAGGAMVAVQAASEV
AARLDVGSITAAVNGPNAVITAGREVERLRIAEFTGLGRTEL PVSHAFSPLEMP
MLEDPRVAGLSYEAPRIALVSNLTGELATEDLVRAEYVWRHRTVRFADGVSL
AAEGASVFLGARDVLTAMAHQHTLDGTATTPALRKDRSEETALLTALALQHVAGVG
VDWMSGIFAGTGRVDLPTYPQFHRYWPRPAVLSGDVWVGAGLSASHPLLGATLLIA
ESGEVILTGRLSLQAYPWLTDHTMGGSVLPFATGFLBLAMRAADQVGCDDRVDFVLT
```

```

PVLVAESGTQVQLVGPADAGRTTISHARPDPDQDPQWQHAGGTLTSGRIYDF
GSTWMPQCAEAVDLEGFYDTTGYGPAFGQLRSVMQCEDGAYVEVALPQVADAGAF
GIHPVLDAVLQSRMAGIGRAGDTFLPFANRGVSLHAGGASVLRARVTKTGEDTLUL
VAVDAEGAPVLSVEBLVRLTQLTHTDAPLTARRGAERDALLSLEWVSGPEVPAGVRC
VSLGADELGVGASVATLADLSCDEDFVVPVSGVGDVPAAAHELTARTLGLLQEWSSG
REGTSARLGVFTRNVAADGGESVVDVAAAVMGVIRSAQSESPCRIALVDLEADSG
DLAEVLAGLPGLLAAGDAQFVVREGAVRVRLVESNEGAGVSPREWDADGTVLVTGGT
CGLGRELARHLVVERGRLHLLVSRGPEASGVGELRDLAAHGAETIAVRACDADRA
AVEVAVVDAADAAHPTAVITHTAGVLDGDIASLTPTERLSAVLRPKVDAANHEATKD
LLOALAPLYSSISGVLGSLQCGNYAAGNVFLDALAGHRAGLGLPAQSLAWGAWTQGGP
MTATLSDADMORMASVGTAPLSPERGLALFDAATAGSATHLVAVGVSVPVRAQGPVP
SMRLGVKGRARTAVNTTGETAVSLADRLRAARPERBTOLMTDVTVEAAAVLGHASS
QVNDARREYELGFSLTSTVEURNRLSTVIGRLKSLSTVVDKSNPTDLAARLYADLAA
QSAIDSPASQAGFLPVAAPETLSRMLFDALDSKVPEAQRMLSALGALRPFENIA
ELEDLPLPATLAEGPTGTPRLICVSTPTANGVHEYARLAASPFGRHSHVSALEPVFAT
GERLPATAVTARVAESALRASDGNPFVLVGHSSAGAFYLAALLENMTWIRPEAV
VLDLTLRHEQONESIDYAGLRMRHFMVDVSPVMTNSRLSAMASWGMGLNOLVVRH
TVVPVLIIRATKETFGITDGTGICYEDHSGSPVDVRSVDADHFSWVRDDAPETAIRVE
WLDLSLGN"
complement(9772..15198)
/gene="pimS3"
complement(9772..15198)
/gene="pimS3"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="pimS3 protein"
/protein_id="CAC20920.1"
/db_xref="GI:12055071"
/db_xref="GOA:Q9EWA2"
/db_xref="SPTREMBL:Q9EWA2"
/translation="MTTSETVVAALRASLKEAENLRQNRKLVAAATEPIALVAMAC
RYPGVQSPEDLWELLTAGKDAIGAFPTDRGWDLDALRDAGVDRRGNAVQEGFLDG
VGDFDADFFGISPRESTMDPQORLLTSTWEAIERAGIDPTTLRSGRTFTGTNGQ
DYAILVRLSADATDGTGIAAATSGRLSYTLGLEGPATVDTTACSSLSVALHTAS
HALRAGECSIALGVGNVMSPGSLMEFSRQGLAADRCCKAFSDTADGTGWSGEGV
LVVERSDARNHGHVLA VRGSAVNQDAGNGLTAPSGPSQOQVIRQALAAAGLSA
DVLVDEBAHGTGTPLDPIIEAQLATYQQDRDPERPLILGSKNLIGHTQOAAAGAAGL
IKMLAMRHGVLPTLYADVPSSHVDWTSGEVELLTPQEPKADRPWRAGVSSFGLS
TGNHVIYVQAEDEEPVPAEPGLAPVAVPMLVSAKTEEAALQARISSRRGTASPD
VGFSLASGRSAFEHRAVLLAQCPDGBLSAARSAETSRPLAVLFSQQAQRVGMGRE
LVYRFPVFAALDEVLAHAGCIQDRVMEGDAEGIDETGTQFALFALEVALFRLES
GVNRPDGHSGISGEIAAHAVGVFLSDACALVAARLMQALPAGGAMVAVQAASE
VVARLDVGSITAAVNGPEAVVLAGDETVARLRIAEALAGGKTKQRLPVSHAFSPLE
PMLEDPRVRAELSYGAPQIPLVSNVTGELAVEELVCSDPYVWRHRTVRFADGVSA
LEAGASFLLELGPDGLVTAMARHTLDETQVAVPLRKDRPEETALLTALHLHAGI
GDWMSGIFAGTGARRVDLPTYPQFHRYWPRPAVLSGDVAGAGLSAHPHLLGATLL
AESGEVILTGRLSLQTHPLTTHAIGNVYFPGTGLELAHAAADQAGCDRVEDLVLM
TPLVLAENGTVQVLLVGPADAGRTTISHARPDPDQDPQWQHAGGTLTSGERVVD
FGSTWMPQDQAVADILEGFGYGPGLRSVMSGCGDGAAYAEVALPDHVAHDAGA
FGLHPALLDAVLQSHRIAGVGSADQFLPFANRGVSLHAGGASVLRVVRVTKDGSWS
IAAVDAEGAPVLSAEAMVWRAVRDPTGTQQAARHGAERDELLSLEWVSGPEVPAGVR
CVSLGADELGVGASVATLADLSCDEDFVLPVSGVGDVPAAAHELTARTLGLLQEWS
GREGSGSRILVFTRNAVSADGESVVDVAAAVMGVIRSAQSESPCRIALVDLEADSG
GDIAEVLAGLPGLLAAGDAQFVVREGAVRVRLVESNEGAGVSPREWDADGTVLVTGG
TGGLGRELARHLVVERGRLHLLVSRGPEASGVGSLDLSLAHGAETVTRACDADR
AAVHALVAAPAEHPHTAVITHTAGVLDGDIASLTPTERLSAVLRPKVDAANHEATK
DLDLAPLYSSISGVLGQANYAAGNVFLDALAHRMGQGLPAGQSLAWGAWTQDG
GMTATLSDTDMORMASVGTAPLSPERGLALFDAACASAAHLVAVGVSVPVRAQGPV
PSVIRGVKGRARTAVNTTGETAVSLADRLRAARPERBTOLMTDVTVEAAAVLGHAS
AKVNDARREYELGFSLTSTVEURNRLSALTQGLRSLAALVDFDTPNPTALADHFPVELV
ERKATDPALLAELERLDSALAGAPDARTAAVTRAAVTRIAQIILDWRDPABEDGAEVAE
RLEAASATDEIFAFIDNELGRHSDR"
complement(15210..43733)
/gene="pimS2"
complement(15210..43733)
/gene="pimS2"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="pimS2 protein"
/protein_id="CAC20921.1"
/db_xref="GI:12055072"
```

```
/db_xref="COA-Q9EWAL"
/db_xref="SPTREMBL:Q9EWAL"
/translation="MDQEQKURDYKIRASADLRSSRQRVGELEAAAEPTAIVGMSCK
PFGVGRPEELWLLAAGDGTGTPSPADRGWELAAAGDGFSGGFLHDAEFDAFFG
ISPREALAMDQOQLLESSEWEAFRAGIDPAAVRGTRTGMFVGAMPQEVGVGPDVV
QGFALTGTTTTFVFNROGLIADGRCSFADSDGTGSGGVGLVLERSEARRNG
AGVTMSPTTFVFNROGLIADGRCSFADSDGTGSGGVGLVLERSEARRNG
HEILAIKGSVAVQDASNGLTAPNGFSQRVIRQALYSARLSADEVDVRAHGTGT
LGDPEAQAALLTYGGDDDDPGLLGSVKNLGHQTAAGMAGVIXMVAHMRGVLPL
RTLHVDAFSDVMDVTPGAVRLLTHEVPMWQGEQFQFVSGTSGNATILIEEAPA
APSPAGGDETPAFAEPAAGVPMVNSRTRREALQAARLLEHLTSRPGGLADJVA
YSLATARGLEHRAFTAAAGDGAARALTALADGTPAPGLVQDARTARTSKLAIFPAG
GSORPMGRELAARPPVATLDELVLHGFDEGLRPLKDVLFPAAGTPEAALLDQTY
AQPALFAEVALYRLAESGTPDPLAGHSIGEIAAHVAGVFTLADAAALVLAGRL
MQALPEGAMVSLATEIEVEVLLLEDRDRLSIAVNGPAAVILAGEADDVTAVEHF
DALGRTKRLRHSHAFSPLEMEPLAEFTAVSRUTPQAPVPPVSGLTGTLATEQRL
TSPDYWLHARHAFADALRLARRRHLPRMGPDGVLSSAMAQSCCRGHRRRRY
LHARRVPRPEAEFTLAAAGLHTRGVQWEPYFOGTGARRADLTPTAFQRRRY
PKSLPATGDDVPAAGLGAHPLLLTAASVANSGLLLTGLRSLRTHPLWADHVRGT
VLLPGTALELAVRAGDNGCGRVEELTAAPLVLPBEGGVQOVWVGPDSAGRRV
SVHSRDPGDEELPQTHAGHVLVAGEHNGFATWNPADATPLDLDGFYARMATGF
GYGPLFQGLRAARWGGDDYAEVLPDSDGATSSSAFGLHPALDORLHAAAFVD
LGEDARGGLPFSQWDTLHAGSASTVRVKLTDPDGDVAVALAVADTTGAPVASIGSLVL

Query Match          7.2%; Score 36; DB 1; Length 84985;
Best Local Similarity 100.0%; Pred. No. 2.6e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 CGGCACGGCGGGTGTGCCCCGACGCGCGCTGCAAG 156
Db 75833 CGGCACGGCGGGTGTGCCCCGACGCGCGCTGCAAG 75798

RESULT 11
LOCUS AX250263/c
DEFINITION Sequence 3 from Patent WO0168867.
ACCESSION AX250263
VERSION AX250263.1 GI:15984065
KEYWORDS
SOURCE Streptomyces cinnamomensis
ORGANISM Streptomyces cinnamomensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Leadlay, P.F., Staunton, J. and Ollivnyk, M.C.
TITLE Polyketides and their synthesis
JOURNAL Patent: WO 0168867-A 3 20-SEP-2001;
Biotica Technology Limited (GB)
FEATURES
source
1. .30000
/organism="Streptomyces cinnamomensis"
/mol_type="unassigned DNA"
/db_xref="taxon:1900"
/note="Nucleotides 60001 - 90000 of the monensin
biosynthetic gene cluster"
ORIGIN
Query Match          6.6%; Score 33; DB 6; Length 30000;
Best Local Similarity 100.0%; Pred. No. 7.9e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 381 GGACGTCACGGGTGGAGGGCGACGGCACCGG 413
Db 23066 GGACGTCACGGGTGGAGGGCGACGGCACCGG 23034

RESULT 12
LOCUS AF440781/c
DEFINITION Streptomyces cinnamomensis polyether antibiotic monensin
ACCESSION AF440781
```

```
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
AUTHORS
FEATURES
source
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

```
AF440781.1 GI:29122977
Streptomyces cinnamomensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
Ollivnyk, M.
The gene cluster for monensin biosynthesis
Thesis (1999) University of Cambridge
2 (bases 1 to 103450)
Ollivnyk, M., Stark, C.B.W., Bhatt, A., Jones, M.A., Hughes-Thomas, Z.A.,
Wilkinson, C., Ollivnyk, Z., Demychuk, Y., Staunton, J. and
Leadlay, P.F.
Analysis of the biosynthetic gene cluster for the polyether
antibiotic monensin in Streptomyces cinnamomensis and evidence for
the role of monB and monC genes in oxidative cyclization
Mol. Microbiol. 49 (5), 1179-1190 (2003)
12940979
3 (bases 1 to 103450)
Ollivnyk, M., Ollivnyk, Z.V. and Leadlay, P.F.
The gene cluster for monensin biosynthesis
Unpublished
4 (bases 1 to 103450)
Ollivnyk, M., Ollivnyk, Z.V. and Leadlay, P.F.
Direct Submission
Submitted (26-OCT-2001) Biochemistry, University of Cambridge, 80
Tennis Court Road, Cambridge CB2 1GA, UK
Location/Qualifiers
1. .103450
/organism="Streptomyces cinnamomensis"
/mol_type="genomic DNA"
/strain="ATCC15413"
/db_xref="ATCC:15413"
/db_xref="taxon:1900"
complement (<1. .1038)
/gene="gdaA"
complement (<1. .1038)
/gene="gdaA"
/note="similar to glutamate dehydrogenase"
/codon_start=1
/transl_table=11
/product="glutamate dehydrogenase"
/protein_id="AAO65782.1"
/db_xref="GI:29122978"
/translation="MTTRPTKTALSKQTSLSQTLTEIHRNPAQEFHQAREVLET
LAPVIAARPEYAEAGLIERLCEPERQIVFVPMQDDHGRVVRNRFVFNESALGPYK
CGLRFPNSVNLGVIKPLGEQIENALTGIGGKGSDPDRGSDAEMVFCQSF
MTELYRHIGHDTVPAGDVGCGREIGYLGQYRRITNRWEAGVLTKGRNWGSLIR
PEATGYGNVLFAMALRRERGETLEGRTAVVVGSGGNVALYTIQKLAALGANAVTCSDS
GYVDEKIDLDLLKQVKEVARVPTVQRRGASARFVFPGRVNEVPADIALPSATQ
NELDADATALI"
complement (1220. .2140)
/gene="dapa"
complement (1220. .2140)
/gene="dapa"
/note="similar to dihydro-dipicolinate synthase"
/codon_start=1
/transl_table=11
/product="dihydro-dipicolinate synthase"
/protein_id="AAO65783.1"
/db_xref="GI:29122979"
/translation="MTLASLEPTPEFLNGLYVPLVPTFDDDLRLALEARLADAEA
LSAGASGLVALGTAAEATLTABERETVIRVCSACRAHAGPLIVGTNDTATAITA
LRELAARGDVAAALVAPPYIRPGEAGTLAHFAALAEHGLPLVVDIPIVPTGTGLTA
GTITAGRLPEVVGIKHATGTSIDPTTWELDSPLPGFAVLGGDDIVLSPLVAAGAHG
IVASANLRTADYAEIMIALWRGSAAPARLADLRLSALFTPEPNTVIKGLVLAQN
RIPSPAVRMPMLLAASADSVRRAPLAASRK"
2211. .3152
/note="SCN 3"
/codon_start=1
/transl_table=11
/product="hypothetical DNA binding protein"
```

/protein\_id="AA065784.1"  
 /db\_xref="GI:29122980"  
 /translation="MLDVRLLHLRLDRRGTTAAVAEALTTTASVSSQOLGLVEREA  
 GVPLLESGRVVLTPAGRLSVHADAVLNKLOAEVAELAGADGGIPLRLITFPFSG  
 GHTIVPALAEIASRHHPALPMVREIDRSVDSGLRAGELDVALVHDYDFVPATPDTT  
 VDEVLLEEMVILVTHAADTATDSGSGSTIAALLGPCAEVPMITARDGTTGHAMAVRA  
 QCAAGFPPIRHHQVDFRTVIALVGAQGGAGFVRMAAEFSPAGVVLTKLPLFRSKV  
 AFRAGGGAHFAIAFAVAANTTAVERMAGSRGPAGGSE"  
 3264..3680  
 /note="SCN 4"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AA065785.1"  
 /db\_xref="GI:29122981"  
 /translation="MADDAYLFLPLDRHPRILGALAAVAGALECTETTPAVHAWLOAHEA  
 SVSSEQVRLPADAEITLIPKDAERLPVLSSEELKVEQECAPQTVTMESEMLAFRE  
 TTQMQALVRLHALTAGIIPAQRIARLTGLDPEIGRL"  
 complement(3684..4307)  
 /note="SCN 5"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AA065786.1"  
 /db\_xref="GI:29122982"  
 /translation="MAVAACAVALVPIDIAVVRISAAVGVLVFFAYLLPYLAILTMTVF  
 VSAPEQVRSWARREARGTFLQRYVLGTAPGGGSLFATAAALVAVLWLPGLHSTTF  
 SALPTLVALLVAAVLCVVAVFVITQADNLVENERALFPPGERSFANADYVYFAL  
 AAMTTFGTVDVDTSRDMERTVAANTVIAFVNTVTVALLVSALGGR"  
 4570..4758  
 /note="SCN 6"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AA065787.1"  
 /db\_xref="GI:29122983"  
 /translation="MTVMDKLQMLKGHEDKAGQIDKAGDFVDGKTQKYGSGQVDTA  
 QDKLRDQFGSDQDEPPQR"  
 5058..5612  
 /note="SCN 7"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AA065788.1"  
 /db\_xref="GI:29122984"  
 /translation="MGTAAQSQEQAAAPGACAEFRVFLCGGGVGLASSFAVVALASWV  
 PWALANALVAVSTVATLHARFTFGAGGATWROHQAAGSAGSAAAYAVTCVAMEVL  
 QQLVAAPGAVLEOVVYLSASLAGVAFVVLRLVFEARNRSLPAAAAVARTARPVRVP  
 APVATVAHASPAGPALCPAA"  
 complement(5693..6010)  
 /note="SCN 8"  
 /codon\_start=1  
 /transl\_table=11  
 /product="acyl carrier protein"  
 /protein\_id="AA065789.1"  
 /db\_xref="GI:29122985"  
 /translation="MTSTDHSTSGQDATELEKQLAAATPBERKLLTTITRTQAGTLIN  
 TTLSDDSNFLGNLSLTALLETTLTGLTGMELAMVAIVENPTPAQLAHLHGGELAH  
 TTA"  
 complement(6045..8531)  
 /note="SCN 9"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ketosynthase-like protein"  
 /protein\_id="AA065790.1"  
 /db\_xref="GI:29122986"  
 /translation="MANEKLVEYLKWTTLAEHLQAQQLRELKAAQHEPIAVVSMACR  
 LPGKTRPDDLDWLVSMEGRVTPDPRANELPEERYALGGLDADAAGDFAGFFD  
 ISDTFAVATEPLQRLMLHAWETVERGHIAHTLKRSTUTGYVGAATHYATLETAP  
 DELLPYLGGSTIAGYSLGLGLEPAISVDTAACSSSLVALHLACQALURGECGL  
 ALAGGQTVMSPTPTFAHOKSLAQDGCKPFAAADGMGLGEGVGLVLERIGDAR  
 KNGHPVLAVIRGSANVQDAGYGLAHPNGQSHVIRAAALDAGITPDQIDAVZAHGT  
 CTRGQKATVONLNTVZAGSDPELNLGTSKNTGTGGAAGAAALIKWVOAFRHG  
 CTRGQKATVONLNTVZAGSDPELNLGTSKNTGTGGAAGAAALIKWVOAFRHG

1 (bases 1 to 20235)  
Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. and Katz, L.  
Modular organization of genes required for complex polyketide  
biosynthesis  
Science 252 (5006), 675-679 (1991)  
91220065  
MEDLINE  
PUBMED  
COMMENT  
FEATURES  
source  
Location/Qualifiers  
1. .20235  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1836"  
19. .20235  
/gene="eryA"  
19. .10722  
/gene="eryA"  
/note="eryA ORF2 encoding modules 3 & 4 for  
6-deoxyerythronolide B formation; putative"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA26494.1"  
/db\_xref="GI:152694"  
/translation="MTDSEKVAEYLRRATLDLRAARQIRELESDFIAIVSMACRLPG  
GVNTPQRLWELLRGGETLSGFPDRGMDLRLHHPDPDNGTSYVDKGGFLDDAAGF  
DAEPFGVSPREAAAMDPOORULLLETSMELVNEGIDPHSLGATGTFPLGKAGYGE  
DTAAAEVGVSVTGAPAVASGRISYTMLEGPSISVDTACSSSLVHLLHVASLGRK  
GESSMAVGGAAVMTAGVDFVDFQRALADGRSKAFAGADGSGFSEGVTLVLLER  
LSEARRNGHEVLAIVRSGALNDGASLPSGPAQRVIRQALESCLGEPFGVDV  
EAGHTGTALDPIEANAALLDVTGDRDADRPLMLGSKVKNIGHTQAAAGVTGLLVKL  
ALRNGELPATLHVPEPTPHVDSGGVALLAGNOPWRGERTRARVSAGFISGNTAH  
VITEAPEERHRETTAHDGVRPLVVSARTTAALRAQAQAIAELLERDADLAVGLG  
LATTAREHRAAVVAATREAEVRLREIAGAATAADAVVGEVTEVDGRNVVFLPQO  
GSOWAGMAELSSSPVPAKIRACDESMAPQMKVSDVLRQAPGLDGRVDPVQP  
VLEAVVSLAELWRSYGVPAAPVQMDLRSVAANGPRSVVSGEPGALRAFSECAAE  
LSGEGMAAVALGAAVPERLRQMDLRSVAANGPRSVVSGEPGALRAFSECAAE  
GIRVDDIDVYASHSPQIERVREELLETGTDIAPRPARTVTHSVESRMDGTEDAR  
YWRNLRETVPADRAVRLASGYDAFIEVSPHVVQAVAEVREAEAGDAGVVGGS  
LHRDGGDLISAFRLSNATVHSGVDLRWDVLPAGPAFPALPTPFORKWYLIQAPAA  
ASDELAVYFWTPIEKPSNGIDGWLVTPLISEPTEMLCEAINANGRGLRCEVD  
TSARTEMAQVAQAGTGRVGLSLLSDGSAACRPGVAGVGLTLVQALGDAGDA  
PWCTLQGNVTPADDDLARPAQTTAHGFAOVAGLEPGRWGVVDLPESVDDAARL  
LVAVRGGRAEDHLAVDRGLRGRVVRASLPQSSRSWTPHGTVLVTGAASPVGDQ  
LVRLADRAERLVLACGPGDGLLAARVEEAGASAVCQDAAALREALGDEPVTLV  
HAGTITNFGSISTEVAPEEPAETIAAKTALLAVDVLGDRVAREVICSVAIGWGA  
GMAAYAGASVYLDALAEHRRARGRCTSVAMTPALPGGAVDDGYLRERGLSLADR  
AMRTWVLAAGPVSAVADWVPLSEGFATPTPALFAELAGRGQAEPDSCPT  
GEPALAGLSPDOENELLEVANAAEVLHESAAEINVRRAFSELGDSINAMAL  
RKLRSASTGLRLPASLIVRHPVTTLAOLHRLARLVGDADQAAVRVVGADSEPIAV  
GIGCRFPFGIGSPQWLRVLBAGANLTTGFPDRGDWIDRLHHPDNDGTSYVDKGG  
FLTDADDFPGFGITPRALAMDPOORLMLTAEVAREAGIDPDALRGDTGTGVPV  
MNGQSYMQLLAGEAEVVDGYQGLNSASVLSGRIATYTFGEFPALTVDACSSLVGI  
HIAOALRGECISALAGGVVMSDPYTFDFSTORGLASDGRCKAFSARADFPALSE  
GYAALVLEPLSARANGHOVLAIVRGSVNDGASGLAAPNGPSQERVIROALASG  
VPAADVVEAHTGTGELGDPTEAGALJATYQDRDRPLRLGSKVKNIGHTQAAAGAA  
GVTKVLAWRHMLPSRLHAEDELPHIDGESGAVEVLRVPHWPAAGERPRRAGVSFG  
VSGTNHAFVVEEAPQEAARTGELPLFVLSGRSEAVVAAQARALAEHLRUTPELGL  
TDAWTLATGRFADRVRAVILGDRAGVCAELDALBGRPSADAPVTSAPRKPVIIV  
FPGQAQWGMARDLWELLESSEVFAESMSRCAEALSPHTDWKLLDVVRGDGDPDHERV  
VLQPLVLSIMVLSAELWRAGVTPAAVGHSGQEIATAAHVAGALSLEAAKVALRSQ  
VLRLEDQGGVSGVARSDELEVTLARMDSGVAAVANGPSTSVAGPTAELDEFFAE  
AEAREMKPRIAVYASHSPSEVARIEDRLAEELGTITAVRGSVPLHSTVTGVIDTSA  
MDASVYRNLRVLPFEQAVRGLVEQGTDFTEVSPHPLLMVAEETAHAGAEVTCV  
PTLRREOSGPHFELRLNLAHVHVGADLRPAVAGRPAPLPTYPFHQRFPVPHRP  
ADVSLAAGEVGLPLVEELVQLRVLVLAGAGALLRMSVGAPDESGRITDVHAAEDVL  
VDLALAGVDRGLVLELVLQPLVLAGAGALLRMSVGAPDESGRITDVHAAEDVL  
DLADAQMSRATGTLAGVAGARDTEQWPPEDAVRIPLDHVDGLAEQGYGSPFO  
ALRAAMKDDSVVAEVSIAADEEGIAFHPVLLDAVAQTLSLIGALGEGGKLPFAMWT  
VTTHASGATSVRVVATPAGADAMALRVTDPAHLLVAITLSLVNRSTGKWEQPEPRGG  
EGELHLDWLRABFGSTGRVVAADASDLDAVLRSGEPEDPAVLYREPEGGDFPRAA  
RHGVLMAALVRRLEQBELPGATLVIATSGATVSDSDSPFEPFGAAMWGVIRCAQA  
ESDPRFVLDTDAEPGMLPAVPDNPQLALRGDDVFPVRLSPLASALTLPAGTQRLVP

GGCAIDSVAFEPAPDVEQLRAGEVRVVRATGVRNFEDVLLALCMYPQKADMCETAG  
VVTAVGPDVDAFAPGVRGLGFCQAFAPATVDRHLLARVPDGSORDAAVPIAYTT  
ADYHDLHLAGLRAGQSVLIHAAAGVGMAAVALRAGAEVLTATAGAKHGTLRALGL  
DDEHASSRETGPAKFRERTGQGVVDVNLSTBELLEDADLAEAGGVFVEMGKTD  
LRDAGPFRGYAPFDLGEAGDDRLGILREVUGLLGAGELDRLPVSAWELGSLVPAALQ  
HMSGRHVGLVLTQAPVDPDGTVLITGCTGTLRLLAEHLVTEHGVHLLLVSRRG  
ADAPGSDLELRAEITDGASAEIAACDTADRDALSALLDGLPRLTGVVHAAGVADGL  
VTSIDPEAVEQVRLAKVDAWNLHETLNTGLSFVILFSSAASVLAGPGGQVYAAANE  
SLNALAALRRTRGLPAKALGMWQAQASEMTSGIDRIARTGVAALPTEKALALFDSA  
LREGGEVVRPLSINRALSRAEFVPLRGMVRKALFAQGAARAGNPNVDRLAGRSE  
SDQVAGLAEVRSNAAVSGVADOLPERKAFKDLGFDSLAAVLELNRLTGTAGVRL  
PSTLVDHPTPLAVAEHLRDLFAASPAVDIGDRDLDELEKALEALSADCHDDVQQL  
ESLLRWNRRADAPSTSAISEDDELFSMLDQRFEGGSDI"  
19. .4470  
/gene="eryA"  
/note="putative"  
/function="approximate span of module 3"  
97. .1482  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoacyl ACP synthase  
of module 3"  
1693. .2670  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyltransferase domain of  
module 3"  
3406. .3921  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoreductase domain  
of module 3, possibly non-functional"  
4171. .4428  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyl carrier domain of  
module 3"  
4471. .10722  
/gene="eryA"  
/note="putative"  
/function="approximate span of module 4"  
4471. .5847  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoacyl ACP synthase  
domain of module 4"  
6054. .7026  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyltransferase domain of  
module 4"  
7165. .9216  
/gene="eryA"  
/note="putative"  
/function="approximate span of dehydratase and  
enoylreductase domains of module 4"  
9433. .9984  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoreductase domain  
of module 4"  
10225. .10483  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyl carrier domain of  
module 4"  
10723. .20235  
/gene="eryA"  
/note="ORF3 encoding modules 5 & 6 for  
6-deoxyerythronolide B formation; putative"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA26495.1"  
CDS

```
/db_xref="GI:152695"
/translation="MSGDNGMTEKRLRYLKRVTYVTELSVTARLEVEHRAEPIAIV
GMACRFPGDVSDFWFSGGDGAIAEAPADRGWEDDPDARLGGMLAAGDFDAGF
FGISPREALNDPQRIIMLEISWEALERAGHDPVLSRTGATGFTGVGTGVIYGRPDE
APDEVLYGTGTGTVASVAGRYAYCLGLEPAMTVDTATSCSGLTALHILAMSLRDEC
GLAAGLVTVMSPGAFTEFPGSLAAGDCKPFKAADQGFGLAEGAGVILVQLISA
ARREGRPVLVRSAGVNDQASGLTAPSPGAQQRVIRRALENAGVRAGDVVYEAH
GTGRLGDPVIEVHALLSYGAERDPPDLWTGSKVNSIGHTTOAAGVAGVAKVALAR
HEMPRTILHDFSPQIENDLGAIVSVQASRWAPAGERPRAGVSSFGISGTAIVIV
EEAPADPEFPDPSGPGVPLVLSGRDEQMRQAQRLADHLAREPNSLRDTGFTLAT
RSAMVHRAVVVQDRLDALAGRAVDIADRTATQARTRRGVAMVPPFGAQQG
MARDLLRSQVFPADISICERAPLHVDMSITDLISGARPLRVDVDPFAVMVWS
AALWESHGVEPAAGVHGSQGAIAAHVAGLVLEDAAKLIVAVRSLVRLALGQGGWAS
FGLGTEQAERLGRFAGLSIASVNGPRSVVAGSGPLDELIAECBAELGRTARIPV
DIASHSPQVESLRRELLTELAGISPVSAADVALYSTTGQPIDTAMDTAYWVNLREQ
VRFQDRLQAEAGDFWVSPHVPVLTGVEIATIDLSALPADAGACVVTGLRDRGGL
ADFTHALGAYAQGVDEVPAPADARVPELVPFQRYWLPITPGGRARDEDDDW
RYQVVRREAESAGSLAGRLVLTGTPVSELSDAIRSGLEQSGATVLTCDVESRSTI
GTALBAADTALSITVLSILSRDGEAVDPSLDALALVQALGAAGVAPLWVLRNAVQV
ADGELVDPQAQMVGGVGRVGLIOPCRMGVLDLVDADAASIRSLAAVLADPRGEQV
AIRADGIKVRLVPPAPARATRWSPRGTVITVGTGGTGAHVARWLAESGAHLVLL
GRGADAPASLREBELTALGTGTITACDVADRAKLEAVLAERAEGRVTSANVHAA
GVSTPTDLDTAEETETADVKRTGNLDELCPDLDAFVLFSSNAGVWVSGFGLASY
AANAFLDCFAARRRSEGAFTVSIAMGLWAGQNMAGDEGEYLRSQLRAMPDRAVE
ELHILTDHSGTSSVVDMDRRFVELFTAAHRPLFDEIAGAAEAQSEEGFALAQ
LAALSTAERREHLAHLIRAEVAVILGHGDAAIDRDRAFDLGLFDSMTAVDLNRLLAA
VTGVRRAATVVDHPTITRLADHYLERLVGAEEAQAPALVREPKDADDPILAVGMA
CRFPFGVHNPGLWEFLVGRGDVATEMTDTRGMDLALDLPDQDRHGTSTYSRHGAFLD
GAADFAAPFGISPREALNDPQRIIMLEISWEALERAGHDPVLSRTGATGFTGVGTG
QYGVQDVPVPEISGVLGTGNSAVVSGRYAYVVLGEGPAMTVDTATSCSLSLVHAG
GSLRDCGLVAGVSVNAGPVEVTEFGRQGLVAGDRCCKAFSAADQGFGLAEGAGV
VILQRLSDARRAGROVLGVVAGSAINQDGSNGLAAPSGVAQQRVIRKAWARAGITGA
DVAVBAHGTGRLGDPVIEVHALLSYGAERDPPDLWTGSKVNSIGHTTOAAGVAGV
IKVVLGNRLGVLPMCLGRSRLIETWSGVSGLAEAVSWPWPAAQVRRAGVSAFGV
SGTNHVTITAEPEPELPEPGVGLAANSPVLLSARTETALAAQRLLESVDD
SVPLTALASALATGRAHPRAALAGLHQLRQGLRAVAGVAAFGAATGTTASAGV
VVFVPGQAGQWCMARGLLSPVFAESIAECDVLSEVAGFSASEVLEQRPDAPSLER
VYVQVPLESVISLRLMAGCGVSPASVIGHQSGEITAAAVAGVLSLEDGVVVALR
AKALRALAGGWSVLSAOPGERARALIAWEDRISVAANSPSVSVSGDPEALAEIV
ARCEDEGVRATKLPVDSVSHRSHVEIRETILADLIGISARRAAIPLVSLHERRDG
ADMGPRVYNLRSQVDFDAVSAADGATFVMSHPVLTAAQVEIADAAVAGS
LHRDTAEHLIAELARAHVGVAVDMWRNFPAPPPVAPLNPFPQRYWLPAPVSDOL
ADSRVYRVDWREPLATTPVLDLGGFLVHGSAPESLTSAVEKAGGRVVPVASDREASAL
```

Query Match 6.4%; Score 32; DB 1; Length 20235;  
Best Local Similarity 100.0%; Pred. No. 0.00024;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 GACGTCGACGGGTGGAGCGGCACGGCACCGG 413  
DB 1006 GACGTCGACGGGTGGAGCGGCACGGCACCGG 1037

RESULT 14  
LOCUS AR049368 20235 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 3 from patent US 5824513.  
ACCESSION AR049368  
VERSION AR049368.1 GI:6005407  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 20235)  
AUTHORS Katz, L., Donadio, S. and McAlpine, J.B.  
TITLE Recombinant DNA method for producing erythromycin analogs  
JOURNAL Patent: US 5824513-A 3 20-OCT-1998;  
FEATURES  
source Location/Qualifiers  
1. .20235  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN

Query Match 6.4%; Score 32; DB 6; Length 20235;  
Best Local Similarity 100.0%; Pred. No. 0.00024;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 GACGTCGACGGGTGGAGCGGCACGGCACCGG 413  
DB 1006 GACGTCGACGGGTGGAGCGGCACGGCACCGG 1037

RESULT 15  
LOCUS AR095529 20235 bp DNA linear PAT 08-SEP-2000  
DEFINITION Sequence 3 from patent US 6004787.  
ACCESSION AR095529  
VERSION AR095529.1 GI:10023473  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 20235)  
AUTHORS Katz, L., Donadio, S. and McAlpine, J.B.  
TITLE Method of directing biosynthesis of specific polyketides  
JOURNAL Patent: US 6004787-A 3 21-DEC-1999;  
FEATURES  
source Location/Qualifiers  
1. .20235  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN

Query Match 6.4%; Score 32; DB 6; Length 20235;  
Best Local Similarity 100.0%; Pred. No. 0.00024;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 GACGTCGACGGGTGGAGCGGCACGGCACCGG 413  
DB 1006 GACGTCGACGGGTGGAGCGGCACGGCACCGG 1037

RESULT 16  
LOCUS SEERVABS 20444 bp DNA linear BCT 28-FEB-1992  
DEFINITION S.erythraea eryA gene for 6-deoxyerythronolide B synthase II & III.  
ACCESSION X62569  
VERSION X62569.1 GI:46977  
KEYWORDS 6-deoxyerythronolide B synthase; acyl carrier protein;  
acyltransferase; beta-ketoacyl synthase; dehydratase; enoyl  
reductase; ketoreductase; multifunctional subunits; thioesterase.  
SOURCE Saccharopolyspora erythraea  
ORGANISM Saccharopolyspora erythraea  
REFERENCE 1 Bevitt, D.J., Cortes, J., Haydock, S.F. and Leadlay, P.F.  
AUTHORS 6-deoxyerythronolide B from S. erythraea: cloning of the structural  
TITLE gene, sequence analysis and inferred domain structure of the  
multifunctional enzyme  
2 (bases 1 to 20444)  
Eur. J. Biochem.  
Bevitt, D.J.  
Direct Submission  
Submitted (19-SEP-1991) D.J. Bevitt, Biochemistry Dept., Univ of  
Cambridge, Tennis Court Rd, Cambridge CB2 1QW, UK  
For related sequences see X56107 & M63677.  
COMMENT Location/Qualifiers  
FEATURES  
1. .20444  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/strain="NRRL 2338"  
/db\_xref="taxon:1836"  
/clone="SE55, DE10, pBK25"  
/clone\_lib="genomic in lambda EMBL3"  
37. .42  
46. .20268

RBS  
gene

[illegible]



```

SSFGMSGTNAH"
ORIGIN
Query Match          6.0%; Score 30; DB 1; Length 1209;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 CGGCGAGTCTCCATGCGCTCGCGGCGG 77
DB 519 CGGCGAGTCTCCATGCGCTCGCGGCGG 548

RESULT 18
LOCUS AF235504 77534 bp DNA linear BCT 19-JUL-2000
DEFINITION Streptomyces hygroscopicus var. ascomyceticus
LDLVMDAGQGVETVCCAGTAAGSGPHEFTRLAGALRGIAVRAVPQPGYEEGEP
SSMAAVALQVADVIRCTQGRFVVGAGSAGALMAYALTELLDGHHPGPGVLLIDV
PQHODAMNMLELTALFRETVMDDTRLTALGAYDRLTGQWRPRETGLPTLLVS
AGEPMGPWDDSKPTFFEHDTVAVPDGHFTVQEHADATARIHIDAMLGGSN"
misc_feature 10867..12111
/gene="eryA"

Query Match          6.4%; Score 32; DB 1; Length 20444;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 GACGTCGACGGTGGAGCGCACGCCGCGG 413
DB 1033 GACGTCGACGGTGGAGCGCACGCCGCGG 1064

RESULT 17
AY283054 1209 bp DNA linear BCT 20-MAY-2003
LOCUS Saccharopolyspora erythraea modular type 1 polyketide synthase
DEFINITION gene, partial cds.
ACCESSION AY283054
VERSION AY283054.1 GI:30961866
SOURCE Saccharopolyspora erythraea
ORGANISM Saccharopolyspora erythraea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.
REFERENCE 1 (bases 1 to 1209)
AUTHORS Nydegger,S.O., Schaeffer,P.M., Born,R., Stamford,N.P.J. and
Haedener,A.
TITLE Coding region of a ketoacyl synthase domain
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1209)
AUTHORS Nydegger,S.O., Schaeffer,P.M., Born,R., Stamford,N.P.J. and
Haedener,A.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2003) Department of Chemistry, University of
Basel, St. Johanns-Ring 19, Basel CH-4056, Switzerland
FEATURES
source
1. 1209
/organism="Saccharopolyspora erythraea"
/mol_type="genomic DNA"
/db_xref="taxon:1836"
<1..>1209
/notes="ketoacyl synthase domain"
/codon_start=1
/transl_table=11
/product="modular type 1 polyketide synthase"
/protein_id="AAP40326.1"
/db_xref="GI:30961867"
/transl_table=11
/translation="VGMACRYPGGVTSPEDLRLLAAGTDAVSAPPADRGWDRDLVP
REGGFLHDAEFDSPVSGISPREALANDPQORLLLEYSWEAFEPAGVPLSVRSGETG
VFAGVMYQVALRAGCEGSMALGGVTVMATPGTFVEFARQQGLASDCKAFKPGADGAT
LVSLHVALRAGCEGSMALGGVTVMATPGTFVEFARQQGLASDCKAFKPGADGAT
GWSGAGMLLVRLSDARRLGHPLVAVRGSANQDGASNLGTPAPNGSPQORVTLQAL
TNARLSTQDVVFAHGTGTALGDPVIAQALISAYGRDRSVPLILGSKVSNLHTQAA
AGVAGVMKMLALRHGVVPTKLHVAEPSSSHVDWSSGAVSLARDLVANPEVDRPRAAV

```

```

SSFGMSGTNAH"
ORIGIN
Query Match          6.0%; Score 30; DB 1; Length 1209;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 CGGCGAGTCTCCATGCGCTCGCGGCGG 77
DB 519 CGGCGAGTCTCCATGCGCTCGCGGCGG 548

RESULT 18
LOCUS AF235504 77534 bp DNA linear BCT 19-JUL-2000
DEFINITION Streptomyces hygroscopicus var. ascomyceticus
LDLVMDAGQGVETVCCAGTAAGSGPHEFTRLAGALRGIAVRAVPQPGYEEGEP
SSMAAVALQVADVIRCTQGRFVVGAGSAGALMAYALTELLDGHHPGPGVLLIDV
PQHODAMNMLELTALFRETVMDDTRLTALGAYDRLTGQWRPRETGLPTLLVS
AGEPMGPWDDSKPTFFEHDTVAVPDGHFTVQEHADATARIHIDAMLGGSN"
misc_feature 10867..12111
/gene="eryA"

Query Match          6.4%; Score 32; DB 1; Length 20444;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 GACGTCGACGGTGGAGCGCACGCCGCGG 413
DB 1033 GACGTCGACGGTGGAGCGCACGCCGCGG 1064

RESULT 17
AY283054 1209 bp DNA linear BCT 20-MAY-2003
LOCUS Saccharopolyspora erythraea modular type 1 polyketide synthase
DEFINITION gene, partial cds.
ACCESSION AY283054
VERSION AY283054.1 GI:30961866
SOURCE Saccharopolyspora erythraea
ORGANISM Saccharopolyspora erythraea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.
REFERENCE 1 (bases 1 to 1209)
AUTHORS Nydegger,S.O., Schaeffer,P.M., Born,R., Stamford,N.P.J. and
Haedener,A.
TITLE Coding region of a ketoacyl synthase domain
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1209)
AUTHORS Nydegger,S.O., Schaeffer,P.M., Born,R., Stamford,N.P.J. and
Haedener,A.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2003) Department of Chemistry, University of
Basel, St. Johanns-Ring 19, Basel CH-4056, Switzerland
FEATURES
source
1. 1209
/organism="Saccharopolyspora erythraea"
/mol_type="genomic DNA"
/db_xref="taxon:1836"
<1..>1209
/notes="ketoacyl synthase domain"
/codon_start=1
/transl_table=11
/product="modular type 1 polyketide synthase"
/protein_id="AAP40326.1"
/db_xref="GI:30961867"
/transl_table=11
/translation="VGMACRYPGGVTSPEDLRLLAAGTDAVSAPPADRGWDRDLVP
REGGFLHDAEFDSPVSGISPREALANDPQORLLLEYSWEAFEPAGVPLSVRSGETG
VFAGVMYQVALRAGCEGSMALGGVTVMATPGTFVEFARQQGLASDCKAFKPGADGAT
LVSLHVALRAGCEGSMALGGVTVMATPGTFVEFARQQGLASDCKAFKPGADGAT
GWSGAGMLLVRLSDARRLGHPLVAVRGSANQDGASNLGTPAPNGSPQORVTLQAL
TNARLSTQDVVFAHGTGTALGDPVIAQALISAYGRDRSVPLILGSKVSNLHTQAA
AGVAGVMKMLALRHGVVPTKLHVAEPSSSHVDWSSGAVSLARDLVANPEVDRPRAAV

```

```
FALHGTQDAQGYADNSGLSFADRYGFLLVFAETTSNNANRCFNMFQSSDNRGOG
EASTROMAATSVAYGADPORVTYITGLSAGAMTSVMLATYVPDVFOAGVAVGLPEG
CATDVSSAYLCNMFGTDLTADQWARRVRDGYPSWSGPMPRVAIHGDKDTIVARNAD
ELRQDQWAVHGVSQTPDRSTVIGNSSTHREYLAADGSVAVEVNRVPGIGHGTVPDPE
TGACQGSTGAYAFIDSTCSSYWTQFFGLSGSADPGSLPAPSLAATGATDTTISL
TWKPVGDATDVAVHRDGAQITTSATTSYTDGLRAGTSHTYAVARADADGKAGLSGA
VTAQTTGATACVWTAGNVAHYQAGRATTAGTYVYAKGSGQNGLYNTEFTVTTLKESPT
GYFTVANDTCP"
complement (3969. .4496)
/gene="fkbk2"
complement (3969. .4496)
/gene="fkbk2"
/note="transcription regulator"
/codon_start=1
/transl_table=11
/product="Fkbr2"
/protein_id="AAF86382.1"
/db_xref="GI:9280384"
/translation="MVEGQGRYFEDFTTGQVYRHLGRVTLETDNSMLTLLTQNTAP
LHFDHYAGTTWGPVLDSTFTLLAVTGQSVTDISQHVMANLGWDRVRLPNPVEGD
TYISQSEVLNTRSSARPDVGLISVRTIGYQDGVIVITFETLNVYRGGHGPFAAV
EPWDRSRRAAGLH"
complement (4595. .5488)
/gene="fkbk1"
complement (4595. .5488)
/gene="fkbk1"
/note="transcription regulator"
/codon_start=1
/transl_table=11
/product="Fkbr1"
/protein_id="AAF86383.1"
/db_xref="GI:9280385"
/translation="WEIRLEWFTALAEHEVTTAERLINSQPTLSRAIRLERRVG
VLFDRHQNLKINKEYEVRFAHRAURALSEVSGAEQRTALIDPDTGTVALGFLHSYG
TWLVPDLGARYRPAFTTFLRGLDAADVDDVNSRDLGLTSPRAGDDVVEVTPFL
RDEALCLVPPGHLIARRRVRTAELADEAFVALEPVGRLQITDRLCAAAGFVPRV
LESTELSTRALVAAGLVAVAPLTGSPSPGDHAVAFLPDPQARRTVGVVTLAGGP
RAPVVARFVDVYRSRAHPV"
5601. .6818
/gene="fkbE"
5601. .6818
/gene="fkbE"
/note="l-carnitine dehydratase"
/codon_start=1
/transl_table=11
/product="FkBE"
/protein_id="AAF86384.1"
/db_xref="GI:9280386"
/translation="MSEPHRPRQERPPAGPLSGLLVVSLEQAAVAAFATRHLDLGAR
VTKIRPGSGDLARGVDRTVRGMSSHFMVLRKESVOLDVRSPEGNRLHALVDRAD
VLVQNLAPGAAGRLASATRSSRGATEADHLRTYPATAVPAATADRKAVDLLVQCEAGL
VSTIGTPTPKVGLSIADICAGMTAYSIGLITALLUKRATIGRGSQLEVSMLGAEWM
GYAEYTRYGTPAPARAGASHATIAPYGFTTDDGQTNILGQNEREMASFCGVLIQR
PGLCDDPRSGNADRVARHTELDVALVSEVTGLTGELVARLEEASIAVARQTVREF
SEHPOLDRGRWAPFDSPVGALEGLIPPTVFHGEHPRGLRVPFELGEHTESVLAIAA
PHSADREEAGHAE"
6808. .8052
/gene="fkbF"
6808. .8052
/gene="fkbF"
/note="similar to Rhizobium trifolii MatC; dicarboxylate
carrier"
/codon_start=1
/transl_table=11
/product="FkbF"
/protein_id="AAF86385.1"
/db_xref="GI:9280387"
/translation="MPNELTGVLLIAAVFLLVAGVGLNMLLALVATFLLGVVALDRT
PDEVLAGFPASMFVLVAVPIFVAFAVRHRTIDPLYAGLMAVNGAAGSAPSGVLFGLA
ALLCAGTAAAPAAVAVPIFVAFAVRHRTIDPLYAGLMAVNGAAGSAPSGVLFGLA
HSALEKNHLFVSGGLFAGTFAPNLAFAVAVSLVLRRLLEPHLDLDDTDTPEGDPAF
RPGAEHWMTLMAAULVTGTTSLDGTFLALTLAALLLPRTSQATKEITAMPVW
LLVCGIVTVYVALLQELGIVDSLCKMIAAIGTPELLAALVICYVGGVVSAPASTTGILGA
```

```
IMLPSEPLFKSGAIGTGMVMAALAAAAATVVDASPFSTNGALVWANAPERLPGVYQGL
LWMGAGVCALAPAAAWAAAFVVA"
8156. .8824
/gene="fkbG"
8156. .8824
/gene="fkbG"
/note="O-methyl transferase"
/codon_start=1
/transl_table=11
/product="FkbG"
/protein_id="AAF86386.1"
/db_xref="GI:9280388"
/translation="MANQITLSDTLAYVRKVSRLRDDRVLRLRAQTAELPGGVLVPV
QAERGQFLFLVRLTQARQVLEIGTYTGYSTLCIARGLAGPGRVVTCDWMPKWEVGE
RYWEEAGVADRIDVRIGDARTVLGLDDEAGAGPESFMVFIDADKAGYPAYEALP
LVRRGLIVDNTLFFGRVADEAVQDDPTVAVREINAAALRDDRDLAAMLTADGVTL
LRKR"
complement (8802. .9884)
/gene="fkbH"
complement (8802. .9884)
/gene="fkbH"
/codon_start=1
/transl_table=11
/product="FkbH"
/protein_id="AAF86387.1"
/db_xref="GI:9280389"
/translation="MTIVKLVLDLNTLWRGTVLEDDDEVVLTDEIRVITLDDRGI
LOAVASKNHDLAWERLERLGAEVFLARIGMGWPKSQSVREIATELNFAPTIIAFID
DQPARAEVAFHLEPVCYPAEQAATLISLPEFSPVSTVDSRRELVMYQAGFARDQA
REAYSGPDEDLRLDLSMTAPAGEEELSVEELTLRTSQNNATGVHYSDADLACF
TDPAREVLVVTMGDRFGPHGAVGIIILEKKPSTWHLKLLATSCRVVSGAGATILNWL
TDQARAGAHVLVADPRFRIDRNRMEIAYRFAGFADSDCPCVSEVAGASAAGVERLHE
PSARPAPTTLTAAADIAPVTVSAG"
complement (9895. .10995)
/gene="Fkbi"
complement (9895. .10995)
/gene="Fkbi"
/note="acyl CoA dehydrogenase"
/codon_start=1
/transl_table=11
/product="Fkbi"
/protein_id="AAF86388.1"
/db_xref="GI:9280390"
/translation="MPERDALITLVGDRAAEWDTSGELPRDLLVRLGADGLLCAEVA
AEHGLGLGSRENGFTAHVSLCSLRVMTSQMAAWTVORLGDAGORATFLKELT
SGKLAAGFSSERQAGSDLSAMETRVRLDGTAVDGHKVTWTTAAAYADHLVVGLED
GSGAVVVFADITPGVVRVERVKPSCRAAGHADLHLDQVVRPAGAVLAGSGASLPLV
AASLAYGRKSVAGCVGIIILACRTAAVAHARTREQFGRPLGDHQLVAGHIAIDLWTRAQ
IAARYCEYASDHWDEGSPMVVPATILAKHVAERAAAGAAATAAQVLIASAGAREGHVVE
RAYRDAKLMETIEGSSSEMCVNLQAHALAPA"
complement (10988. .11248)
/gene="fkbj"
complement (10988. .11248)
/gene="fkbj"
/note="ACP"
/codon_start=1
/transl_table=11
/product="Fkbj"
/protein_id="AAF86389.1"
/db_xref="GI:9280391"
/translation="MTAPVKDDPVAAHALVAFILKTKTRSDWPFVDRDLFAEGGLTSLFAM
ELVVVLEKTFDVTQIAGPDQLQLANFRFTVESMWALVHRLRAVDA"
complement (11245. .12093)
/gene="fkbK"
complement (11245. .12093)
/gene="fkbK"
/note="acyl CoA dehydrogenase"
/codon_start=1
```

Query Match 5.8%; Score 29; DB 1; Length 77534;  
Best Local Similarity 100.0%; Pred. No. 0.0058;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





2935..*21081	VTGGDLAQSAVCLVRSQTEPNDRIVLVETSSDAHAGAGSGVDDADLAEVLVPLGIIA	Query Match	5.8%;	Score 29;	DB 1;	Length 78210;
/gene="pteal"	SSEPQAAHADDVVRPRLVRVAVPDAEPHELPGSGSGTLLITGASGGGLGLPARRHVA	Best local Similarity	100.0%;	Pred. No. 0.0057;		
/codon_start=1	EHGVRSLVSRRGSDAPGAELTADITAGADVTAAACDVADAAVRAALLAGPGQSL	Matches 29;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
/transl_table=11	SAIITHAGVLDGGIIGSLTPELTDVFRPKVDAAALNHLDELAGELSGLSAEVLFSVSA					
/product="modular polyketide synthase"	GTGTGQOANYAANTFLDALAEHRRAGORATSLAWGLWAQDSSEAMTGLTDLT					
/protein_id="BAB69303.1"	RTIIGTGLVLDGGIIGSLTPELTDVFRPKVDAAALNHLDELAGELSGLSAEVLFSVSA					
/db_xref="GI:15824139"	RTIIGTGLVLDGGIIGSLTPELTDVFRPKVDAAALNHLDELAGELSGLSAEVLFSVSA					
/translation="MVSESNATGNRDCSASADSGADGIAVIGLACRLPAGADPAAFWR	ARRTQAPAPSSDGLLFDLALADRAVAPSLDGLGQGTVPVPLRGURGT					
	GFDSLTAEVLNRLNSACVGMRLPATLIDFYNPLTIAHLITGEAAGAAATVPTTSL					
	TATGTADPELTAIVGMACRYFGVQSPEDLWOLFSGRDAVSGFPEGGRWDVYIDPN					
	PDWGTSTVTEGGLFHDAADPAEAFGGISPREALAMDQQRLLETWEAFERAGIDP					
	ATVRSQTVTEGGLFHDAADPAEAFGGISPREALAMDQQRLLETWEAFERAGIDP					
	TVDYACSSSLVALHQAALSGSCSNALGVTVMGVSVFVFSRQGLSAGRCCK					
	AFGQATGTWAGGVGMLVERLDARNGHVLAVRGTAIVNODGASNGLTAPNGPS					
	LKSNIGHQAAGGVGKVMWMMQHVLPQTLHVDEPTPHVDWSAGAVQLTETAEVW					
	PASDGPRAAASVFGVSGTNAHVIQEATSEVTARTAGVAVGPVWVLSAKSDAALRE					
	QAGRLISFLDAEGVGGAGVRPEDVAYSLATRAVLECRAAVVEGLFEGFQGLEGIS					
	GAAVPGVAGVGRVGLFSGQSORIGMRELYAAVVPFAAAVDEVCALLDATTVDAAE					
	TLHOTGSTQPALFAVEVALFRLLSSWGRIPDYVAGHSGETAAAHVAGVLSIEDAAKL					
	VSAARAALMQALPAGAMVAIQATEAEVLPHLTHGVGLAAINGPSPVLSGEDAVLDI					
	AEVFAAQGRKTSRLKVSHPAFSLPMDPMLFEFAEVVGLTFSEPPRIPVVSNTLGRIAE					
	VVIPEYVVRHVREAVRFADGLTHLHGVTFVEIPGQVLSALVQCGCLDGGTHITVTV					
	PALDRAPEPVALTAVGQHLTHGISPEWQALFPGARRVDLPTTAFQERXYLWDAPOA					
	SGDVRAAGLGAADHPLCGAATATADSGAVLTGLSLGTHPLWADHANGVILLPCTA					
	FVELAIRAGDETGCWITIEDLLELPLVYPERGGVTLQVAVGPDESGRQLSVHGRVD					
	DRSMIRHATGVLSSAAEAFSAKIGAWPPAGAEVDLTAFYEGMAAAGLDYGYFPQGLR					
	SVWRDGDVFAEVELPEDESAFAAFAALHALLADALHAAAGVLSLDDSPALPFAW					
	SGVYVQAAGAAVMRVKLSRTATGVTAVADGAGDPVASVLSLRASAEOUPAGSRETA					
	GDQLPGLTWTFTLPSATDALMIETVADFALRDESEGTDDVVVVVPPCPAGSRETA					
	DRVHSTSEVTLALVQMLAEFRAGLAVTRPGLAHAAVWGLVRSVSAOSENDRITLV					
	ERHETDEAVRVLPAALASGEPOFAYRDEVEFVPRVVKATRSATGATGTFDGVGVL					
	LTGASGALGVLVRLHVAHGVSRLLLSRRGAEAPGAVELAEALAAWGAERVMAACD					
	VADRAEAGMLNGLGEQSLSAVVHTAGVLGDIVASLTPERMREVRFPKVDVNLNHE					
	CTRDMLAALVFFSVAGVAGSAGQAAAYANFLDAFSAHRRRGLPAISLAWGWE					
	ERAPVLRALVPAARATARAAPASASSIAERLEEMPEAREKTVLDIRVRAEAGL					
	GHASDTRVPEHAFODLSDLSITAVELNRNLKATCLRPLATLVPHPTPLARHVF					
	MAAAGAEPALVDSLLADLQVEQELVTLTAASEARDRLSKLQGLLALAGFEKGAPA					
	QRPGETGGSDLESATDDEVDFLLGKEFGIS"					
21273..*31847						
/gene="pteA2"						
21273..*31847						
/gene="pteA2"						
/codon_start=1						
/transl_table=11						
/product="modular polyketide synthase"						
/protein_id="BAB69304.1"						
/db_xref="GI:15824140"						
/translation="MSCRYPGGVSDPEDLWELVAGGRDAVSEPPADRGWDLDALYDPO						
	PSRSTYAREGAFURDLARFDAGLFGISPREALAMDQQRLLETWEAFERAGIDP					
	ASLAGSRTGVFAGVWNNHEFLAALQNSPDLEGLGTGSGSVASGRVATYFGLGEPAV					
	TVDYACSSSLVALHQAALSGSCSNALGVTVMGVSVFVFSRQGLSAGRCCK					
	PRAAGVAGTGWGSGAGLLVERLSDARRNGHVLAVRGTAIVNODGASNGLTAPNGPS					
	QQRVIRQALAAQLTAAQVDAVEAHGTGTRLGDPTEBAQALLATYQGRPDKPLTWLGS					
	IKSNIGHTQAAAGVAGIKVMWMMQHVLPQTLHVDEPTPHVDWSAGAVQLTETAEVW					
	SASDQPRRAGTSFGISGTNAHTIIEEAPSAEFAEQDTGAPPEGFEAGAESAVSEA					
	ARPAAGPFWVTSRTEALRAQAEELRTSDAEVPPDLDAFSLATRTALAEHRAVL					
	QGTPTPAEAGLADAVVDSBELHRTGAOPALFAVEVALFRLLSSWGRIPDYVAGHS					
	AFAAVDECAQILDADVDDVDSBELHRTGAOPALFAVEVALFRLLSSWGRIPDYVAGHS					
	SVGEIAAAHVAGVLSLDDAKLVSARLMOALPAGAMVAQVATDEVDLPHLTDAYG					
	IAINPTFRPQIVPVVNSLNTAGVTPQYVHVHREAVFADGVKRTLNDLGVTFVEIG					
	PGCVLSALAEAGCLDDVVTVPALRADRPKVALVAVSELHAHGVSPDWHAFPPGAR					
	VDLPTVAFQYERYWLDASPAAPGGVRAAGIGSADHPILGAAVSLAGDGERLLTLGRSL					
Qy	382	GACGTCCACCGGTGGAGCGCACGCCAC	410			
Db	47502	GACGTCCACCGGTGGAGCGCACGCCAC	47530			

```
RESULT 22
AF357202      113193 bp      DNA      linear      BCT 04-NOV-2002
LOCUS      Streptomyces nodosus amphotericin biosynthetic gene cluster,
DEFINITION      complete sequence.
ACCESSION      AF357202
VERSION      AF357202.1  GI:14794889
KEYWORDS
SOURCE
ORGANISM      Streptomyces nodosus
                Streptomyces nodosus
                Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
                Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE      1 (bases 1 to 113193)
AUTHORS      Caffrey, P., Lynch, S., Flood, E., Finnan, S. and O'Liinnk, M.
TITLE      Amphotericin biosynthesis in Streptomyces nodosus: deductions from
                analysis of polyketide synthase and late genes
JOURNAL      Chem. Biol. 8 (7), 713-723 (2001)
MEDLINE      21344785
PubMed      11451671
REFERENCE      2 (bases 1 to 113193)
AUTHORS      Caffrey, P., Lynch, S., Flood, E., Finnan, S. M. and O'Liinnk, M.
TITLE      Direct Submission
JOURNAL      Submitted (07-MAR-2001) Industrial Microbiology, University College
                Dublin, Belfield, Dublin, Ireland
FEATURES
source
1. .113193
   /organism="Streptomyces nodosus"
   /mol_type="genomic DNA"
   /db_xref="taxon:40318"
   complement(4..1824)
   /gene="amphg"
   complement(4..1824)
   /gene="amphg"
   /note="possible amphotericin export protein"
   /codon_start=1
   /transl_table=11
   /product="AmphG"
   /protein_id="AAK73498.1"
   /db_xref="GI:14794890"
   /translation="MAPDVEEHEEERESEQPVRRLLAALRRPHRRSVGLALTAVGVGI
                LNAAPVLLRLLHDLRAOEKLARLPLRVDFRKPACETISRGTDVNDLQOQLT
                WLAVSVLLHDLRAOEKLARLPLRVDFRKPACETISRGTDVNDLQOQLT
                DLISVSVLLHDLRAOEKLARLPLRVDFRKPACETISRGTDVNDLQOQLT
                LRAHEMCAHALVAFDRRAEORFERNREAVTRAGSQAQFAPGAEFVPMFYAN
                LGYVAVVGAMKVNGLTGDVQAFILYARQFQPIVEIASVAGRLQSGVASARV
                FTLLDAPEQEPDFRPLAVRVEFRVQFVSFRYSPDPLIBGLSLVPEPGSTAVV
                GPSGAKTVALNLMRYEIDSGRIILDDTDAAMNDDILSRFGLVQDTWLFKGTI
                AENIAGSGATRADIVAAATYADRFTLTSQGYDTLDDSGGVSAGEKQLITVA
                RAFLARPAVLDEATSSVDTRTELLIQRAMNLRPQRTSFVIAHRLSTIRDAVIV
                MKSGRIVEQTHQDLDAQAYARLHARADAADAVTVG"
   complement(1805..3628)
   /gene="amphH"
   complement(1805..3628)
   /gene="amphH"
   /note="possible amphotericin export protein"
   /codon_start=1
   /transl_table=11
   /product="AmphH"
   /protein_id="AAK73499.1"
   /db_xref="GI:14794891"
   /translation="MARSVKRPFPQRAAGRPQCHSVLLRLMETQLRPYAGSVVA
                LVVHLVQLLGLTLLTGAAALIDEGVRRHSDRIGTIGATMAVVALVQIAALGAA
                LGARTSALGRDLRSVAFRVLDPSAREIGRFGPSSLTRSVNDVQVQMLAAGGIGI
                FVAAPLCLGSLIALAQDTLAILVPMVLVAVCFGLLSIRMAALVQLQOGLDRI
                GRILRLIYGVVRVSFARDAHEGFRTRNEELLSGLVGRILVAVLPSVILMMNL
                FTLLQVLTGVRVSRQAGALSAFLSYLSLIIMSVMLAFVFLNVPRAVCAERTI
                EVLQELTDDVPPASPRMAGPAGQVELVGHAFRFPYAGNAVRLSLTLRPERVAVL
                GSTSGKTLHLHLILVTVAGVIRGGITDRLDPSVMAAAVGYVQFRLPAGV
                ASNNRFRPATDELMEVLRVIAQHEGFVTLRGLDTEIAQGGTVTSGGQRQRIAR
                ALLRRPAIYLPDSSFDALDQSTEAARKALVPYEGATVITVAQRVASVRDRAVILL
                DOGGIATGTHDALLRDSPTYREIATLSQRTREETAHGAGRS"
   3840..4874
gene
```

```

/ gene="amphDIII"
3840..4874
/ gene="amphDIII"
/ note="mycosamine biosynthesis protein"
/ codon_start=1
/ transl_table=11
/ product="AmphDIII"
/ protein_id="AAK73500.1"
/ db_xref="GI:14794892"
/ translation="MPKRALITGITGQDSYLAELHLLSQGVQVWGLIRGQAMPKPRFV
                SRLASLSFVGDMLDQSLVAVDQVDEVYVNLGALSFVWMSQOAEVTVANGMG
                VIRVLEAIRVWSGLSMSTAGTEQIRFYQSSSEMFCVARETLPHEPRSPYG
                AAKAYGHFITRNYRESFGYVAGMLFNHESFRQGEFVTRKISLAVARIKLGLODKL
                ALGNMDAVRDMGYAGDYVAMHMLMQDAPDDYVIGTGMHTVRDAVAFRHHGLDW
                KQYVVVDQLVRAEVEVLCDSSKAQQLGKPSVDFQELMRMVDADLASVSQNE
                LDDILLAHSM"
5042..33574
/ gene="amphI"
5042..33574
/ gene="amphI"
/ note="polyketide synthase multienzyme polypeptide housing
                extension modules 9, 10, 11, 12, 13, and 14"
/ codon_start=1
/ transl_table=11
/ product="AmphI"
/ protein_id="AAK73501.1"
/ db_xref="GI:14794893"
/ translation="MDNEKRLDYKLATADLRARRRRVGELESQSEPIAVGMTCR
                YPGVSSPEDLRWVYBENGVTPTPTDRGMDLEALASSPTSGRGFLHDPADDF
                GISPREAVMDPQQRVLESWEAFERAGINTSVKSGTGVFIQMAQDYVGPADG
                AEGFQLTGNTGSLGRISYTFGTGPVAVTDTACSSILVAVHLATQALRGCTIAL
                AGGVTVMSGPTFIEMGRQGLSVDRGCRSGFTADTGAGVGLVILERSLDAIRN
                GREILAVRGTVANODGAGGLTAPNGPQQAVIEQALNARLSAGDIDVEAHGTGT
                TIGDPVEAQALLATYQQRDEKPLILGSKNISHTOAAAGVAGVKNVAMRHGVL
                PRILADHPTRVNDSQGRVLTENTWPAAGPRAAVSSFGISVNAHTIVEOAP
                EPEADPEDTTPAATVGVLPVLLSGSRSEVLRQAALLTLTGTPPPADLAV
                SIATTRTAFHRAVLIALDPELTGRITAIAGTDPAVLDTVGTARTIRLAVLT
                GOGAQLAGRELAAFPVFAAALDAALDAFTPLDPLRVKLVGMDADRLDTEVAQ
                PALFAVEVALYRLLESFVKPHLAGHSVGEIAAAHVAGVSLDDAATLVAARLMO
                ALPEGMAVAVQASDEVAFLAGHEDVLAANVGPASVLSGDDTIVTELAAALAA
                DOKTRSLVSHAFSFLMAPMLDFRNVVEGLTLHSLPLVSDVGTGPRATVOLT
                PIVYWDVQAVRFADGIDWLKHDVITAFELGDFSDVSLNMAQNCIDAGSDALTVA
                LREGRDEHTTAAALHTQGTALHWDACFTGTGARTDLPITYTQRRYWPRAVQG
                GAADLRSGVLAHHPLLSAAVLSADSEGLATGRISLHSHWLDHVTGVALLPBT
                AFLEAVRAGDEVGCRVDELTLAAPLVPEQGVQVQVQVQVQVQVQVQVQVQVQV
                DADEDAPTSHTAVLSASRTSDFDATVMPADAETLPVGLYERLAEFGGFGYGLF
                QGLRAWRGDEVFAVLPESGHTDAESGHLHPALLDSALHAASVLDLRAAGGLP
                FSWEGVSLHSGATTLVRILAPAGDAVAIADDSGQLVLSADSLIRAVAREIDA
                AAALVRDALFRLDVPTVAASGTAAALVGEDPFGRLALPQSGDLAHFPLDLAAA
                DGAVPTVLLPTGTGPDADPVTAAHRAATEALAAVRLTEQDERFAASRLAVTRGA
                TTGHDPAALAVMGLVRSAGSGRFLVLDLDDQDTPALPAAALTSSEPOLAVRGE
                LRAALVRPASTAEAVPAFGGEGAVLTGCTGLGALVLAHLVABHGVRELVLVSR
                GGAAGAAELVAELAESGARATVACDVTDRAAVELAAHPVSAVHVSAGVLDDGV
                GLTPRELITVLRPKVDAWNHLEATRDLDLKAFVLPSSVAGVLSGPGQANYAGNAF
                LDLAARRRAAGPLGLSLAWPMEGTGGTGGISEDDLRMARAGTAPALIVEQGLALL
                DAALDGDAAALAFVRLDLSVIRAQGVPPPLRLSLIRGRSRAAVAGATAGLAQRLL
                RLDAKRDDELVLVQVQALVGHATGAIEDAGRAFRELFGDLSLVALELRLNTVT
                GLRLPATLVFDYPTVSHLSYVLDLGLTEVEAEVVGQRTAAVADDPIVTVMACRYP
                GGVTSPEDLWRLVTEGTDVSGFPVNRGVDENLYHPDPDHPGTAYTRSGFLHEAGE
                FDFPFQMSPREALATDSQORLLHEASWEALERAGIDPVGRLSGATGVFAGVMSYS
                AMLSGFEFGFGGSGSSPLASGRVSYTLGLEPFAVTVDTCSSSLVAMHMAQALS
                GETSLAAGGVTVMSVPAFVDFARQGLSPDGRCKAFSDSADGVSGVMGLVLR
                QSDAIRNGHOILAVVRGSAVNODASNGITAPNGSQQRVIRIQALASGGLSAGDVV
                EAHGTGTLGDPLEAQLATYGRDRDPROPLILGSKNISHTOAAAGVAGVKNV
                SMHGVLPTLHVDASSSHVDVTEGAVELLTEQATPETGRPRAAVSSFGISCTNVH
                TVLSQAPGTTPFAAAPTAVAGVPLLSGRDTRDALRQAPVLAHLTHLQHPESLADL
                GHSLATRPFERRAAVIAQDREGLASGLAAGRPDPAVVEGAAGRAVAMFGSG
                QSORAAMGRRELYETOPFAADEFVCAALDPLDLRLPREVFAAEGSEEAALDRTG
                WTQPALFVAVLYRVISWGRADFVTHSIGHIAAAHAGVTFLQDAARLVAARAT
                LMAALPGGAMVAVQATEEVAPLLGRGLSVAVNGPSTSVVSGDELPVAAAEFSG
                RGRYRRLVRSHAFSPHMDAMLDLDFVTAETLSYAPRILPVSDDLGRADDAEVT
                AVLRVRRVRAVAFADCVTRLDAGATFLDELGDTGLLTAAEATLGDERTHTALV
                PLLRADRPEESAAATAARLIQIHGVLDLWTAYLAGTGARRVLDLFTYAFQHAHYWQUP
```

CDS

gene

CDS

gene

SAAPSAGDPAQKNAVERGDAEELAAVLGLDESLTPLDLSLLPALSSWRRGNQEK  
ALDUTKRYVWTRLSKPAAPVLDGTWLLVSSDATADDTELLDGLAELGARGVR  
RLVLDACQRAVLGRLADENADNTAQVLSVLPDERIDGTDPAGFTQGLALITAIIV  
QALADTGAGRLWATRGAVSTGPAFDPVTHPAQTAWMGVGHLEPRUWGLGVLDLP  
ADPDGAGQRLAVGQAPDCEQVALTRGVGRVLRHVHIDELPSQATGASG  
VLITGTGGGLAGTARSLARAAHVLVTSRGPDPAGAAELAEFGASVSVIAC  
DVADRALAVLGLSDADPLTGVVHTAGVGHYGLDALTAEFAGLTAKLAGAAHL  
DNLGDBELFFILFSGIAGVWGSQDSAYGAANAYLDALARRAGLAAWTIACWP  
WGTGWAADDAVSTGRLGLGLDPPALTEMRRAVRQDVTTVADVADWTYAPLF  
TSRPSALTDLPEVRAALAEENTPADTDAEISEIVRVSSEPEQLKLLDILVTEKA  
TVLCHSAGAVLPEDRAFPRETFGSLTAVERLKHGATGLSLPTWVTFDPTLELAQ  
YLRAMVGSVLEAVGVPATGCTDEPTAILGMSCKYPGVSSPEQLMDLVLSGTDAIT  
DFVNRGNTAGLVDPDHPHTYTGQGFLEADEDFPMFGLISPREALVMDPOOR  
LLETWEAFERAGLTPDTRLSLTGTGSSVOEYGMGAGDGAELHVTGTSPLVS  
GRLAYVPLGSGVLTACSSLVLAHLCAQLRNGESNLAVAGATVMTTENAFVA  
FSRQALAQGRCAKSFESADGMTLAEVGVILVERLSDARRNGHVPVLAIRGSAIQ  
DGNASGLSAPNGSQSVIRQALANARVAFCEIDLEAHGTGTPDIPRAQALFAY  
GRTRPTETALLIGSVKNI GHSSQAGVASTIKMVALRHGVMPQTHADPESSHVDW  
SPGTVELLGNTPOTGRPRRAVSFSGISGTHVHILQEETAPAEDEQLAPL  
FVAGVVPILLSARGAALREQADRLTHVATDPAARPIDIGLSLATSRLPHRAV  
VVPAGTDPLELRAVADGSGVVGAVADVAGRTVFPVPGQSQWAGCAQLLDS  
SPVEDLHDELELLEAVPRTSIPFSTVTGDLDTVMNDAGVYRSRGLRVLSAA  
AVRDLIAADHRAFLVSSHVPLANSVQDMIDDAVAGVASCGLDRDGLRFLLSAA  
EVFVGVQVDAVAFEGTGASRVLDPTAFQHENLWMAAAPEAVTAADPEDAFWTA  
VEDGDSALTAALGTDBDSVAAPLSPALSSWRARKERSTVDSWRYRTPKRWTKLPQR  
TLDCGTWLLSVAGDVTDAFALGTGAERVLRLVDESDCTRAVIRLTDADGLTGI  
VSVLGAETGATGVLGVALTVALVQAGDAGIDTPLMALTRGAVSTGSDKVT  
APVQAVTGICWTALECPGWGVVDLPETLDARAGORLAALVLAGLGDQDQALRS  
SGVTRIRVADAPDSARDKPRGTTTLATGSSGTAPHLARLAEQGAHVLVSR  
RPEAPGAERLEAERGETTLAQCIDTRDAVALLLESLEKRGTRVTVKHTAT  
IELEHTLTDLDDFRLAAVTAQILDELLDDELDFFVLSSTAGWGSAAHAY  
VAGNAYLAALAEHRRAGLTALSWSGIVADDLQIGRVDPMIRSGLEFMDPOLAS  
GLKALDDDEQIVADVDMETHPVTYSARPPLFDEVEVQRLTAAEAGSAGPAR  
GEFAALLAPAEHQRLKLTVRTEAASVLGSSAEDLTQDRAVDVFGSLTAVGL  
RNRLASVTGLTLPSTWPDYENPALAGFLHSELADVHSAAGVATAGAPVDDPIAI  
VMSCRYPGGTSQMLWRVSLSEVDVAVFPADRGWDAEALYDPPDASGRITSVQG  
GFLRDVADFPFGGISPREALSDMPQRLLETAWEFENAGLDPVAGRSRTGTFI  
GASYQDYAAVPSGSGSEHMITGSLPSVLSGRVSYLFLGEPVTLTACSSLVAI  
HLACQSLRNGESTLALAGASIMSTPMSFIGFSRORALAEGRCKAYAGAGDGLTAE  
GVGLILRELSDARENGEVLAIRGSAVNOGASNLTPNPGSOQRIQALANAG  
VEANDIVLEGHTGTALGDLPIEAQALPATYKORDPERPVLIGSVKNI GHTQWASG  
VASIILKHALREGVAPKSLHIDQSTHVDMSSGTIQLLTERTEPWTGRPRRAVSS  
FGLSGTNHTVLEQAPADAFAAEDTPAPRDALVPVLSVGRGEALRAQAALLLLA  
ERPGHTPTDLAFSLATSRAALEHRAAVADDHEALVRLGLTALRDLGLPAGLVQGRTR  
GRTPFLTGQSGRLMGRELYRHPVFAADALDAVLRIIDGTTERPLRDLVFAEBSQ  
DAALLHRTGYAQPALFALEVALFRLESNGVTPDYLAGHSVGEIAAAHVAGVLDLDA  
CTLVAARGLLQALPEGGMWVALEAEDEVLPHLEGLADOVSAVAVNGPSVYVAGEE  
EPVLAAAHFAEQGRKTRKLVSHAFHPLMDPLDDPAVAARALTYHAPSIPFVSNV  
TGTTLAEPQVCTDPAVSHVRSVAFADGIMLSTGGVQVTFLELGPGLVLSGMARES  
LTDASRTALLPTLRDREPEQALVAVAAHAGFDVDWTFWFGSGARRVALPTAF  
QSRWYPTDTAGITAPAPGALDAEFWAAVHADVLSITSLGLDDATVTVAMVALLT  
ANWRGRGSLADSNRYRVTKPRGCGAAPTGRWLVLVPAEHRDEATAAADAEEA  
ALATATRVETVTBRALAAALRETAADGTFQGVLSLILAPGAGHGGAALTLT  
ATLQALQAGIARIDALPWNITRGAVAVGRSEQVTAPEQAAMWGLFRAMALLELPARYGGS

Query Match 5.8%; Score 29; DB 1; Length 113193;  
Best Local Similarity 100.0%; Pred. No. 0.0055;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GACGGACCGGCTGTCGAGGCGTCGG 203  
|||||  
DB 51354 GACGGACCGGCTGTCGAGGCGTCGG 51382

RESULT 23  
AX703543  
LOCUS 113193 bp DNA linear PAT 03-APR-2003  
DEFINITION Sequence 1 from Patent WO02097082.  
ACCESSION AX703543  
VERSION AX703543.1 GI:29538472

Streptomyces nodosus  
Streptomyces nodosus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1  
Caffrey, J.P.  
Engineered biosynthesis of novel polyenes  
Patent: WO 02097082-A 1 05-DEC-2002;  
UNIVERSITY COLLEGE DUBLIN (IE)  
LOCATION/Qualifiers  
1. 113193  
/organism="Streptomyces nodosus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:40318"  
5.8%; Score 29; DB 6; Length 113193;  
Best Local Similarity 100.0%; Pred. No. 0.0055;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 175 GACGGACCGGCTGTCGAGGCGTCGG 203  
|||||  
DB 51354 GACGGACCGGCTGTCGAGGCGTCGG 51382

ORIGIN

Query Match 5.8%; Score 29; DB 6; Length 113193;  
Best Local Similarity 100.0%; Pred. No. 0.0055;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 175 GACGGACCGGCTGTCGAGGCGTCGG 203  
|||||  
DB 51354 GACGGACCGGCTGTCGAGGCGTCGG 51382

RESULT 24

AX703543  
LOCUS 138203 bp DNA linear BCT 06-DEC-2003  
DEFINITION Streptomyces sp. FR-008 heptane macrolide complex synthesis gene cluster, complete sequence.  
ACCESSION AX703543  
VERSION AX703543.1 GI:34766435  
KEYWORDS Streptomyces sp. FR-008  
Streptomyces sp. FR-008  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 138203)  
Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K.J., Lee, S.Y. and Deng, Z.  
Organizational and Mutational Analysis of a Complete  
FR-008/Candidin Gene Cluster Encoding a Structurally Related  
Polyene Complex  
Chem. Biol. 10 (11), 1065-1076 (2003)  
14652074  
2 (bases 1 to 138203)  
Chen, S., Huang, X., Zhou, X., He, J., Bai, L., Jeong, K.J., Lee, S.Y. and Deng, Z.  
Direct Submission  
Submitted (30-MAY-2003) Bio-X Life Science Research Center,  
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,  
China

FEATURES

Location/Qualifiers  
1. 138203  
/organism="Streptomyces sp. FR-008"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:206662"  
complement(574..1950)  
/gene="fscQ"  
complement(574..1950)  
/gene="fscQ"  
/note="putative FAD-dependent monooxygenase"  
/codon\_start=1  
/transl\_table=11  
/product="fscQ"  
protein\_id="AAQ82549.1"  
/gb\_xref="GI:34766435"  
/translation="MPSTCAPTSTIGRCPHRRKRGTRGRRIAVPTPCGKVRKGH  
QAKKRWVSRMGSSVAVVGGSTAGCAALAAASRGAERLTVFERADARLDKRGVIGV  
HDAKVAELDAGYMADEMFWALNRRVWSVRDGAHGAIGTGPFPFAYNNGSLWS  
ELRRVPETATYRAGAKVAEQDADGVTVRLADGPEFEDIVIGADYSVVRVYAMY  
PQTGAAYAGYIGWRGTSPDVSGLPSDGLDAHNITFPFGHGMAYRIPDGGGHRNLNWL

gene

CDS

```

YTAPEIRGLHDLRTPTSLPPGRINAELETHRALVAEHPPFPFWAAKLLSTPAETTF
IQPIYLDYPHYADRMVILGDAASAPHLGAGSVKALODATALEAAWAGESWKEI
LEGTHAARGPVTAVUGLARRNGSAQVEETEDWSAQAEFDAMWQEQNGSDRRSGF
GGHSLKSR"
gene      complement (2264. .3037)
CDS       /gene="pabC"
          /gene="pabC"
          /note="putative 4-amino-4-deoxychorismate lyase"
          /codon_start=1
          /transl_table=11
          /product="PabC"
          /protein_id="AAQ82550.1"
          /db_xref="GI:34766437"
          /translation="MIEDGEPAGPEALASLALTYGHFTTLLVNGSVRGDLHLER
LIRDCRTLLDALDPRVRKRLRAAPTDGRATVTVFDPALNIGNIAADARPGLIV
TSRAPDPPGGLVRSVRDLPEVKSGLCTPLRLRQORAGYDDVLTGPDGD
ILEGTMVNGVLRDGEVVMVPGGEVLATGTQLRRGRHRSRHTGRPRGPRLSRGLRD
QRVGVPVVTGIDDEFPAAHSHSVTLAEIYQALPGSL"
gene      complement (3150. .3818)
CDS       /gene="fscRI"
          /gene="fscRI"
          /note="putative transcriptional activator"
          /codon_start=1
          /transl_table=11
          /product="FscRI"
          /protein_id="AAQ82551.1"
          /db_xref="GI:34766438"
          /translation="MDPAPADPAGSPERAAYVADGFDAGAYVACLDPSLTQQVNO
BEPFGSSSLCGRFCDLVPSVQOPLMHOFARMLDKGRHATEVIAVDQRTA
STPLNALAVRGRTPDVAAILVMNAEEEGADADWAPRKKLSEIDARILEGIAA
GVSTPLASRLYLSRQGVHYHTGLRKLKVPNRALVSRAYSIMGVLKGTWPPKIVE
DFIK"
gene      complement (4377. .7205)
CDS       /gene="fscRII"
          /gene="fscRII"
          /note="putative transcriptional regulator"
          /codon_start=1
          /transl_table=11
          /product="FscRII"
          /protein_id="AAQ82552.1"
          /db_xref="GI:34766439"
          /translation="MLERTTIALVDALGAATGGRSSILVLTGTLGIGRALLQRL
SLCADDPRVIRAHAAEMEQDFAPGVVRQLFETLLGDSPEDARDWMDHASFARH
VLADDAAPGDAQIAATEAVLHGLSLIANYADSRLILVDDLQMSDVPSLRWLTY
LAKRLHGLAVVVCALRDGDPDRSHHTLAREIREAGTQTLRPASLSLTAHELVRHFH
EAGDDEFVQACHEASVGNPLFLLSILVGTGFLGRPLASHAETARLRPSQLRERLAS
ILRTQAPVPRDLAAIAILGEQSDAPTLARLAGLDSIGVAGALRALGALGAAPDEP
RTHRSVRDAESTLTWQERHEDPAALLYAAGSPAQVAAQLMAVVTTRQPAVD
VIRAADTLRRGAPTAAGYLCRALDPSPAAGVGRGLLVELGTAEGRFDPACERH
IAQMTLLPEPRDVAALRIISPTALGPAPLTAVDLLRQAEDLGPADGLTGTDRDLA
LRLAELRCHDEDHELAAHRLKELGPEPPVDTGSRKELTIALIGHAITGLCHLPA
SEVRLAGLRDREPATSAIRHSTLPLAIIITMAADSVRAVDSWLAVEHRANERGTPP
ANALVIEQALVHLGRGLAQAREQSETALGTEPAHQGLDTAATHTALVALESDDP
SLAEKVRASRIKRPGLTGLTGLLEATEATAGDGTGALDLSHSCGRLEAGWH
NPLVFPWRPWAYVHRHGDITARSLEAEEYARAEQWAPGVGSGRALLGRLEDGG
RIGLIRGAVTVLRGSANOLELARSILAEIRGGAETEARLAREATLATS CGAFGL
ARRAGAGHRGPPATPDVLTPTTERRVVGLACRGLTNOETAGALGVSRAVEKHLT
HAYRLGISRELIVLFSRAEYAGTD"
gene      complement (7210. .10254)
CDS       /gene="fscRIII"
          /gene="fscRIII"
          /note="putative transcriptional regulator"
          /codon_start=1
          /transl_table=11
          /product="FscRIII"
          /protein_id="AAQ82553.1"
          /db_xref="GI:34766440"
          /translation="MYDLHMPPIELPLERDRELAALSAVIGELSGRPAVVTGEP
GIGQNDLLRRAAYADAGLIVLSAHATPAEHVRVYGVVAQLLAENRALAPLFLTD

```

```

EQPGLGCLINGLNAVRDRPTLLVVVRDQWLDPASRLWEALARELPAPVALLTSTT
GTALTRPMSVGTPLTGLATTIELALPILTSOGTATVLRAPGAPGDPFAFTALAEAT
RGIAPVHDVLDPRFARAGSVPRADRIDPLRALTAEVGVGHATRALSDLRDPAADALR
ALAVCGDILLDFPCTLAGPHSVESRIDLRAALASGLTTRDGRHPRVODAVVRARVLE
EMPAADRAELYARAGLAQORVAADDQGLADLILLARFVGDWPWAVTLRGFTSALRGG
RDLAVAYALALDEPLAEDRARTIEFQLASVEMVTPATAERRRIGGLIRATRPGPGA
GLRARATDCLLGGDTRAAHALAGAIIDSAAPAEPRRGTGAPVPEAEPASTPT
HLVPHRPVAGTPPGATPGDRRSAPDLGTPTRAHDAPRSFGPEAGRPDTPAGPDEDP
AVAEHELTVLCRVASFIRHDDADIDVLPVLPALPDVHNPAVAGVWAEHGVAGVHRE
EVRRLARSAPVPAATGRPLLYVPLLGGCRALLTDDGDAENHIGALLDESHERAAI
SVAHILTVAELHTRHGRPDAAADLAAQAELPLDLRLHPLFLPFLALSMITLQNG
HTRARETAARPLPLAHESATTAQLFARGVLARTDDEPHOARHFRACGRGHLRHG
CANPAVQWRSLAAEAHITLGTTEAARLVHEEVARLAWGAAAPLGRGAASLASLVTE
ENRVNLAAVTTLSAASARTAYTRAVIELAAAREENERRTAVAGLGAATVLPVPAE
PPGHVPASAVARPPGPRPAVLTRRTAKAGHDAAVWGLGTAVERTETALAQAQGLNR
ELATELAVTTAVELRLSGVRKLRIGREELRALVQEAGS"
gene      complement (10298. .13315)
CDS       /gene="fscRIV"
          /gene="fscRIV"
          /note="putative transcriptional regulator"
          /codon_start=1
          /transl_table=11
          /product="FscRIV"
          /protein_id="AAQ82554.1"
          /db_xref="GI:34766441"
          /translation="MWLRWAGDRSRGAAPGRSGSAGPPSLAPSDARESVPHRSRAVP
PAPPLVGRAHQDLDVLARHADARACRPLVLDDGAGICKTALLRAALAECDGCLAGMT
TLYGSCRAVDVATGSGVRALFGGLGTGRKGTSPLLVGGARRALPALAADPGELDA
DAHRTGTQGLTAPDLAANLMDGPIVLVDDVHWCDBERLWLDLFLRRADGLPLLVV
AAGSTGFTLGLTADLANIVAHHLPASLGLPDTAEABLTASHFPDPLPSFVGRLL
LSVTGSPLEIVLLRLDLRSAGLPGDDTGTARLAQAGKVAAVSGVLEHDPVWRQ
VARAVLIGDEDETVLALGRVSVVHVEAVELRLDAGLHGHLELSHDLVRAAIVQD
AVGASGVILRRRAARLSDVGRSPEEINATQLLYPGTDPDWAHSVLRDAAQAQORG
ACEAAARYLDERVREAPKDPDVLISRLKALAEIDPARSVILLHEASLITTDVRAAAT
AVQGLGLTCLAVQSPDGRALTEALDLDALDELPPEPADRLRTLTESALLIVGSD
KATPLILRTTEGLTPQDGTQAQQLAMLSVLAAEGGAEETVGRARRALRAGV
PIGVMSLLPTSLALSIDANEAAEVLTVLRSGDTPAAVTVLALSTRSLFRIENG
AYPDAMADAQTALEILGOERWGDTPMHTAYASVLTERRGDPGRALADAIKRPHL
RFVWYHTYLMARGALAADGDLGALQVFGCGSMAEAGLTNPVLAIPWLETAULL
GEAGREAGRAAAGHTRLAERWGTTRALGAALARGAAGARTGARTGAUREAVALLAD
SPARSCHARGLLGLGALVDGAVREGRHLREAVGLARRCCGVALARRARDELIAAG
GMRREVTASPLDMLTGTERTVAGLVASGAGNREAVESLFTVTRTVELHLTSYVRKLG
AERGDITLAREAGATARPAEERRPGHAKRNP"
gene      13522. .14898
CDS       /gene="fscMI"
          /gene="fscMI"
          /note="glycosyltransferase responsible for mycosamine
          attachment"
          /codon_start=1
          /transl_table=11
          /product="FscMI"
          /protein_id="AAQ82555.1"
          /db_xref="GI:34766442"
          /translation="MDSARPITLVSLPSGLNPLLVLAGELARGVDPDLWPAIDEB
RREVEALSDVSKVSFASGEVVPSEL SAVTDDVEVYVTSQSRFKAHRAVVRVSRP
ALQAKRYLEQVDEVPALMVVDCVAGVDLALARGIPVTVNVFVAVSNVLTSHN
PFGATYTPKSPFPVNSGLSPARMSQKANTLTFKMTLGMFLHDPMAALLREDDAIRK
EGCIAPPNAPNTVDSAAAVVCSVAELDPFDIPDRVSLVAGVLPPLPEAPDDEVTR
WLDAAQSVVYMGFGTITLRTREVAALVEVARRMSGTHQFKLKEQOHLLEAGSL
PDLAVESVPSQDLVLAHPNVSVFPHSGGNAHYEGVFGKVPQVVRPLWVDFCDQAV
RGRDGIISLTLKPHTVDDDDVDKLTIVTSPAPFTEARLGAALQRAAGRAAADL
VTGLLUPAA"
gene      14953. .16011
CDS       /gene="fscMII"
          /gene="fscMII"
          /note="putative aminotransferase responsible for
          mycosamine biosynthesis"
          /codon_start=1
          /transl_table=11
          /product="FscMII"

```



```

/protein_id="AA082556.1"
/db_xref="GI:34766443"
/translation="MATHFVSPALDGRLEVYSDAVSGNLSQGFYVRRERFA
EWNGVAGVACSGTAALTALRALNIPGVGEIVPEFTWVASAWATYTGAPVFD
CGDLNDIVTRIIEKIITPRVAVMPVHYGRCDMDAMDALQYNLRVGEASABG
VRPVDIACTACIKITAGEGVCITDDPRLAOLAHLRAMAFTRDHGFLHKKLAY
NYRTMAGVALAQTREILAREIEIRDAIGKLPGLITMPARDVLYMWDLR
AEREERLHLDARGIETLFFKPMSPQGVLPDVPWTLNAHRSFEDGLYPLTHGLT
AA09EITGAVRDPFRAG"

Query Match      5.8%; Score 29; DB 1; Length 138203;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 CGACGGCCGCTGCAGGCGTTCGCCGACG 169
Db 116232 CGACGGCCGCTGCAGGCGTTCGCCGACG 116204

RESULT 25
LOCUS AP005022 300425 bp DNA linear BCT 10-MAY-2003
DEFINITION Streptomyces avermitilis genomic DNA, complete genome, section
2/30.
ACCESSION AP005022 BA000030
VERSION AP005022.1 GI:29603894
KEYWORDS Streptomyces avermitilis MA-4680
SOURCE Streptomyces avermitilis MA-4680
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE
1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
21477403
MEDLINE 11572948
PUBMED
REFERENCE
2 Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H.,
Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.
Complete genome sequence and comparative analysis of the industrial
microorganism Streptomyces avermitilis
Nat. Biotechnol. 21 (5), 526-531 (2003)
22608306
MEDLINE 12692562
PUBMED
REFERENCE
3 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kushida, N., Director-General of Biotechnology Center, Shiba, T.,
Sakaki, Y. and Hattori, M.
Direct Submission
Submitted (29-MAR-2002) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:bioente.go.jp, URL:htp://www.bio.nite.go.jp/,
Tel:81-3-3481-1933, Fax:81-3-3481-8424)
This work was done in collaboration with Haruo Ikeda (*1), Jun
Ishikawa (*2), Akiharu Hanamoto (*3), Chigusa Takahashi (*3), Mayumi
Shinose (*3), Hiroshi Horikawa (*4), Hidekazu Nakazawa (*4), Tomomi
Osonoe (*4), Norihiro Kushida (*4), Hisashi Kikuchi (*4), Tadayoshi
Shiba (*5), Yoshiyuki Sakaki (*6, *7), Masahira Hattori (*1, *7)
and Satoshi Omura (*1, *3).
Final finishing process and all annotation were done by H. Ikeda
and J. Ishikawa
*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University

```

\*6 Institute of Medical Science, University of Tokyo  
 \*7 RIKEN, Genomic Sciences Center  
 Following url is also available.  
 http://avermitilis.ls.kitasato-u.ac.jp.

FEATURES  
 source  
 1. 300425  
 /organism="Streptomyces avermitilis MA-4680"  
 /mol\_type="genomic DNA"  
 /strain="MA-4680"  
 /db\_xref="taxon:227882"  
 /note="This strain is also named as strain: ATCC 31267,  
 NCIMS 12804 or NRRL 8165.-synonym: Streptomyces  
 avermectinus"  
 complement(382. .645)  
 /note="SAV258"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67967.1"  
 /db\_xref="GI:29603895"  
 /translation="MMNFALAPEVSVTDTEHTGYLLDERTGYWLLNGTGALALRALI  
 DGTSPGVASLQCFPDAGTAQVEQVDNLLHALRTAKVAAS"  
 complement(642. .2552)  
 /note="SAV259"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative membrane-associated Zn-dependent  
 protease"  
 /protein\_id="BAC67968.1"  
 /db\_xref="GI:29603896"  
 /translation="MPAVRLPHLLTIVEGTGMEFLVLPDPAGAAALAMPTAGPPP  
 THVIEHSGREWIVGCVNEEDVMVAVGHRVALLGRTTADADVLTRLRDVRSDRL  
 DGLAATAGSFHVASVNGEVVQGTLSACQIYYATVAGVTLAARFQALADRIGAD  
 IDEQVALILSPFGPPWPLSGVSCPTGCTGCIIGRDGTGTQWMTAPD  
 RPLNGMRLQALAEAVATCTAGORVWSADISGLDSTSLCFADADRHDAHLTVHY  
 QALGRSSDRTWQAQEAALPGARHLVPGQADPWYAPLAHAVHDSGLPTFVSRA  
 TTEHLARLAAGSRHHQIGIGIDELFPSPMGALHVRHPFLAALPHLRALRSRRW  
 PLTTTARTLASAPSTQWLAAACSNFFADRDHWGADWEPLKLPFWATPHAAATVRR  
 QLLAALATPQPLATLPVQHEMIRLIRNGNATRENHIGARTVTFHAPFTDDRVLE  
 AILAVLRDRVALGQFKMLRSIVPAGLLSRTTKGDASAEIYAGLRERRGELLK  
 LMDSHLYRMGFVDADVLRVTLRGLHADTRPLMLPLDSTILACELWRLTSLSPADTARATA  
 VVPSPRIVKEI"  
 3174. .3572  
 /note="SAV260"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative transposase"  
 /protein\_id="BAC67969.1"  
 /db\_xref="GI:29603897"  
 /translation="MGRAGVAFVLTDEQRNRYGTENQVDDQQLAGYFLLDRDRRA  
 MTCGARSQLGAYVQLGTVRFLGTFLDNPEDAPAEVYAVYAEQLGHPGSLVAGYGAER  
 TWMDHQIAIKDAYGYRDLKGDAAWKLKRDGD"  
 3604. .3834  
 /note="SAV261"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67970.1"  
 /db\_xref="GI:29603898"  
 /translation="MLGQQYVPLDLMIEGESQRMSEDRALRGLRMWVIAIEVPV  
 YLVWVVEVSLTGQQQFGRTEEDGKPRADTD"  
 4027. .4500  
 /note="SAV262"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67971.1"  
 /db\_xref="GI:29603899"  
 /translation="MOADAMAVAVMELQGAANRLWEPAGRGRTFLTLTLLAFAGG  
 AARARTAGTQVQSLGVFGQAAELLSRRVASKOTVAAREPLAQAAATAAPLMRCS  
 APAVVATQQLNLAADVNTARLEALEAFGRVNVATPPRLSWARRAVSSG"  
 complement(4811. .6445)  
 /note="IS6 family; completely same as SAV7545 and SAV75480"

SAV263"  
/codon\_start=1  
/transl\_table=11  
/product="putative Is6 family transposase"  
/protein\_id="BAC67972.1"  
/db\_xref="GI:29603900"  
/translation="MILARCGELLKTFPHLEGLVLEVEPEGGVLRHIVARTVESVVP  
FCDPATPSVRISGYORRLADGAVGROVSIETVRLPCDDORCARVFAEVDGL  
TVRGRTPOLRLLSIAI VALAGRAGERLAARLPVSRTRTLLALWALPDYAEPT  
RVLGDFRATRGKHYGTIVLCETHAPIDLLPVSATFAAMITDHPVEIICDRRG  
GAPADGATGAPAVQVADLWHLWHLNLAETVKSLSKSSCLREPALASESPPEVLRP  
PLSHAGRLARARHAAVHDLGGLTIRAVARRLSLRSNTVRYARVATPPEELATG  
RWQNLPSLDPYKYLHQRWHEGHTSGAKLHAELERGFTGYSYVRYLRRFRRTSG  
DRPPAPPPGVKVTHTDRMDNDQDKLAILARCELEAATCHVRSFAAMMA  
IRRSARLPEWATARADEHGLRGFADGLLADLDAVILGLSTEWSSGCGVEGRVTDIKL  
LKRQAGRAGLPLLRKRVLLVAADRQRHVTNQTAH"  
/notes="SAV264"  
/complement (6592. .8802)  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC67973.1"  
/db\_xref="GI:29603901"  
/translation="WTPATPTPPSTPPSPWPCQGGTTPEDPAGCRGIRVSGHTACLA  
HLSADRDAYLAGSPGNVDRHGTFSQDILLDALRDPDGTGDPRLGVARFRSAT  
FQDVGFRSATQKWADPGSATFQGNAGFRLATFEGDAGFSAREGPDAGFSARFEG  
DASPVSAITFEGDASFSATFQGNAGFRLATFEGDAGFSAREGPDAGFSARFEG  
FGADFRGADFRGADFRGADFRGADFRGADFRGADFRGADFRGADFRGADFRGAD  
ATFQSDTLEAATFSDHTFDRATFPGDARFVEATFRTGTFVSATLQSNANFLGVT  
ERDAKFESI FQGDATFVSATFECDAFCEQGVFERSASGLPLVCAGRVKLSGAVFGCP  
VTLSPATCLCRFRWSTAEALRLRYATVDFAHAVFEVPLTIAEAADPFVLHNGRPV  
AEQSLIDAPDIEVRIASLRGVDAAHLVLSVDVLSGLCTGTGTHLDQIRLEGTCSPASV  
PRRTHWRRLPFRFQRALAEZGFWRASQPTFPGVNAAVAGAHAGPAOLAPVYRA  
LRKAFEDRNEFGAATFYGEMEMRHDTGSTRTEGRLGGLLGLSLGLASRALG  
WLAAMLVITVLLMFGPLKSPKQVATGTPPGDKVTLEIDKADPQNTGDRFTSE  
RFEKALTGTLSNVSQDQLTTAGTITEMASRVTEPVLGLGLAVLRNVRK"  
9294. .9707  
/note="SAV265"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC67974.1"  
/db\_xref="GI:29603902"  
/translation="WGRILRLRADPSCRCRKLTKQVPSRPGGVQGLITNTVLSLP  
TVEYDLDARFTGGLRVTSRQBLGLWLAEGYLRPQ"  
complement (9955. .10614)  
/note="SAV266"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC67975.1"  
/db\_xref="GI:29603903"  
/translation="MPHVLRRLRESSGVYINLRADANRLPVQFPIILLHNRHRRD  
DNRPGWTEILRLSALGRTAEASHALPAVAGEPAENRLOSLSGQAVELVRP  
ACRCMRARFGHVLVKTTPPEAVCLRHORWLLGBQYRLHALPELOQANREHCRIV  
RRTASILAREYAVARDCLURDFQDGKPSALRQWNRKIRLDQLEDPEFTGDTVPVAPPSP  
G"  
10830. .11180  
/note="SAV267"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC67976.1"  
/db\_xref="GI:29603904"  
/translation="MLEDRAGVLDLCTWVHLTGDWPATRAHAEFVFLPHLRSH  
RVRLVQVARASRSLEITVIDDSRQPIRGVPGFWALLWDEYSGGTVPGQGIKLSLH  
AKGNWRPLSGTTC"  
complement (11530. .11979)  
/note="SAV268"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"

/protein\_id="BAC67977.1"  
/db\_xref="GI:29603905"  
Query Match 5.8%; Score 29; DB 1; Length 300425;  
Best Local Similarity 100.0%; Pred. No. 0.005;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 382 GACGTCGACGCGGTGGAGCGGCACGGCAC 410  
Db 216856 GACGTCGACGCGGTGGAGCGGCACGGCAC 216828  
|||||  
RESULT 26  
AF453501/c 82746 bp DNA linear BCT 18-JUN-2002  
LOCUS AF453501  
DEFINITION Actinosynnema pretiosum subsp. auranticum maytansinoid antitumor  
agent ansamitocin biosynthetic gene cluster I, partial sequence.  
ACCESSION AF453501  
VERSION AF453501.1 GI:21449342  
KEYWORDS  
SOURCE Actinosynnema pretiosum subsp. auranticum  
ORGANISM Actinosynnema pretiosum subsp. auranticum  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Actinosynnemataceae; Actinosynnema.  
REFERENCE 1 (bases 1 to 82746)  
AUTHORS Yu, T.-W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K.O.,  
Xu, J., Moss, S.J., Leistner, E. and Floss, H.G.  
TITLE The biosynthetic gene cluster of the maytansinoid antitumor agent  
ansamitocin from Actinosynnema pretiosum  
Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7968-7973 (2002)  
JOURNAL 12060743  
MEDLINE 22056096  
PUBMED 12060743  
REFERENCE 2 (bases 1 to 82746)  
AUTHORS Yu, T.-W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K.O.,  
Xu, J., Moss, S.J., Leistner, E. and Floss, H.G.  
TITLE Direct Submission  
JOURNAL Submitted (28-NOV-2001) Department of Chemistry, University of  
Washington, Box 351700, Seattle, WA 98195-1700, USA  
FEATURES  
Location/Qualifiers  
1. 82746  
/organism="Actinosynnema pretiosum subsp. auranticum"  
/mol\_type="genomic DNA"  
/strain="ATCC 31565"  
/sub\_species="auranticum"  
/db\_xref="ATCC:31565"  
/db\_xref="taxon:42198"  
misc\_feature <1..82653  
/note="cluster I of the maytansinoid antitumor agent  
ansamitocin biosynthetic gene cluster"  
gene <1..1678  
/gene="asm37"  
CDS <1..1678  
/gene="asm37"  
/codon\_start=2  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAM54115.1"  
/db\_xref="GI:21449383"  
/translation="IYPLGLVLAATGVLAAAPAAITSPVTSTATVTFEVRGTGT  
PAPSRVYVDRSAATGYAQGGGADHLLNGSLYRHSGGGWATQVRYPTFSASGV  
ARWTVRDLGEGASPNVDVDFQVEAPVETSAKVTHVTYTAGHSSVAYQPSANVFN  
PERGLYHTGDCDKADLSADALRYRTQEQITLNCVFYLAERSSPIATDLRLQR  
QFDAARVAGVQLVRLFAVYATASEGSDAPKERVLAHIDQLAPLIRANADVTAVVQAGFV  
GANGEVYTKNFGNAGVSATQANRKAIVVDKLLTALPADRAVQLRTPKFKALYGGG  
STRIHNDCLFLASDITDFYEDPAPYPLEAETRLPMGGTCAVNPRTDCTPAT  
AEISRHWYTLNTDIYQDVLKWSAQCLAEVTKLGRYTLRDSVPEPTTTRGAALP  
RVNLTNDGVAAYNQRTNVLNRTTGLTLPADPRTWLPGRARAQDAQIIP  
SLATGOYALLLELADPALPGRPEYSIQNTANTGLNDARTGLNDKQTIVS"  
1825. .2142  
/gene="asm36"  
1825. .2142  
/gene="asm36"  
/codon\_start=1

```

/transl_table=11
/product="unknown"
/protein_id="AA054111.1"
/db_xref="GI:21449382"
/translation="MVTASAVLCARAGCAPAGATARWCFSTACKQAAVRLRRKA
GASHEHEDEQLREALALVARATILTEATPWSAQATLVGLARRVWELAAADGAV
SGS"
gene
complement (2129..3307)
CDS
/gene="asm35"
complement (2129..3307)
/gene="asm35"
/codon_start=1
/transl_table=11
/product="bicyclomycin resistant protein"
/protein_id="AA054113.1"
/db_xref="GI:21449381"
/translation="MAVGKADSAWGVLPAALLIVALTGTITIDMSVPFDDIARLGL
VGPAVQLVTNLFGLGICGALCHGFLSDRFGRRPVLGGSAVMLVGLGALCAPTIG
LILTARFVOGIGASAAVLFVLIIGDRYGAALRMVGLVNAASALMTCAPLGGAI
NRVAGRGNYLAVCATLSLLAVFLFELPSTRTRATITAVSVLADYRLILGSRAYL
ASLTSLLPAAVLFVIAASPELTYAAGLGVGTGEALHLLVVVSAFVSAPATGILP
LUGGPAGAVRWGUGAVVVGVLGGLGDPVATVGSVALPCVGFACVYFVVFERSV
FPLAGASATNGLRALLVTLTLAGALSTGGPFAPAAVMAGAAVAVVLLSALSFR
"
gene
complement (3450..4064)
CDS
/gene="asm34"
complement (3450..4064)
/gene="asm34"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AA054112.1"
/db_xref="GI:21449380"
/translation="MARNPERRAALDAAVEVLAGEGARGCLTFRAVDQAGVPAFTA
SNYPAGAEILTAGERYVQRLDQAFIGETPDGPRDRSRVTELMHALVDRVSAFPTG
FLALLELRLEARNPRELVEITAFVRADLRFNAHHAASGLPGDGTVALILNALNL
ILRLTLPLDILTHQRDLVTALVDRLLAGQEPAGSPFDESGR"
4151..4723
/gene="asm33"
4151..4723
/gene="asm33"
/codon_start=1
/transl_table=11
/product="dihydrofolate reductase"
/protein_id="AA054111.1"
/db_xref="GI:21449379"
/translation="MRKLIVFIASIDGVIAAGPGEHDFYTAEPWVHLRAEYPETT
PTHVPLVGMPLDTPNRYDTVLMGRGSPALDAGVGSFYAHLRQYVVSRLSPDADP
AVALVRADPVLGVLRELKREDGLDVLGCGADLAGQLLPEIDELVVKTYPVVAGGTPL
FRAGFAPRFPDVRAPFDHGGVVTYRPA"
complement (4854..5057)
/gene="asm32"
complement (4854..5057)
/gene="asm32"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA054110.1"
/db_xref="GI:21449378"
/translation="MWITQONARIIPARRSGRSNARRCAVRWPRQPRSATRPFGRS
SANADSRARSGRDPRTSILNRS"
5415..6461
/gene="asm31"
5415..6461
/gene="asm31"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor"
/protein_id="AA054109.1"
/db_xref="GI:21449377"
/translation="MWIGYSADGRRRRFRVWGVTEAFMTWETDEDIKLAATLRDAEA
ANAASTRAYLKAIGRVLLTADQEVLEARVVEAGLYAEHLDDLPLPPARGSRPVELR
RVVRDQQRKDHLEALNLRVLSIAKRYTGKGMALDVIQEGNLGLVRAVEKDYAKG

```

```

FKFSTYATWIRQALTRALADOSRIPLRVHNVVTVTKLARLERDLHNGREATAE
VAVELDVTPKEKVEVLRQHARTVPSLDQTIAGADTSIGLIIADENAVAPLEAVSSALL
SADLRVLEITPPREAGVVRMYRGPFDGVLHTLDEVGRVYGVTRERIKQIESKIMAKL
QHPARSGLLEGYLD"
gene
complement (6501..9518)
CDS
/gene="asm30"
complement (6501..9518)
/gene="asm30"
/codon_start=1
/transl_table=11
/product="cytochrome P450"
/protein_id="AA054108.1"
/db_xref="GI:21449376"
/translation="MVATGTRIPGPKPLVGNLLDVLTSDDLTDVDFLDRCHREHGG
IVALTPAQORQVFASHELVFMSWGAHVPALEQVRDFPAGGLTFARGDEFNW
GKAHRLMFAFQPTAMRHFPAMLDIAEQMLVRRFRFPDHRIDVADDMTRGLDITIA
LCAFGAFNSFYRDRAPFVDMVRSLVEAGERAERLPGVQFLVGNRQYRDDIATM
NRADGIVARAALPAGERDDLLERMLTCADPTGTERLSARNRVQLQVLETRLHP
TAPAFALAPDEPAELGCGHAIGAGEPVLVMLPTLHRDPVAVWRDPDVPFERFAERWDE
IPACAMPPFGHARACIGRPFALQEAATLVLLVLRQFDLALADPDHRLTIKQITLKP
DSLIVRARPRADPGATATVETVYVHQVPAHRRGTPLHVFGNSGSGGLARTIAG
DGAARWATSVAPLDDAVRALPASGPVIVSSSYNGAPPDNAAHFVRWLTQDGLDLSG
VDYLVLCGNLWSATYQVPTLIDEMAAGARLRERGATDARADFFGDMWRWYEP
LWELLSAECGVEGEIGPFRFVESDAADGLGDLASLVLENRELVRGPDAGSKRHE
LRPDGTSYNTGDLVLPONHEDLVRRVAVARLGTAEARVTVVESAPTCGLVPVGRAL
RVDELITRCVDLSPAGAGVVARLAEPCPPERAEALANTGAILLELRFERSCAVD
LALALELPAPTRLYSISSAAEQRAEVALTVSVTGSYLSRVRPDKRVAGTAS
PPESFRPADNTVPVLLIAGTGIAFPGRFLARALAGGEPGALLFGCGPELDDL
YAEFAALGDWLEVDVRAHSDPGEVHRVQHLRQRRDRVRELVDAGARVYLCGDATR
VGPAVEVLGRIGFGAGWLDALRAGGYATDVE"
9664..10245
/gene="asm29"
9664..10245
/gene="asm29"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AA054107.1"
/db_xref="GI:21449375"
/translation="MSAARREQLVAAALVRMKDGIATAATTRAICAEADMPHGAHY
CFRSQBELYALTGTSLDAAPASISDLPVAVNVRTMLAAFTAFQADPRAELV
LGEVLTLAQREPELRELPAHQREYVAKAVAHLDRLAEAEAGFDYSVDRLAELVVS
LDGVASSWFSRDRGDLAKASLDQFVAMFAALTR"
10393..10857
/gene="asm28"
10393..10857
/gene="asm28"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA054106.1"
/db_xref="GI:21449374"
/translation="MTDTTSHRTPVONAAALVAGVFLVLLGIAGFVPGLTNTYDGLF
AGHSDDLGLGVFKVKSALHNLVHLARGAGTAAGLSKAPRGARFLIGGGAIVAVLWL
YGVVVDHSTANFVPLNADUNLHGLVAGVMIGLGAALGRNAHVRAARSDD"
complement (10912..11937)
/gene="asm27"
complement (10912..11937)
/gene="asm27"
/codon_start=1
/transl_table=11
/product="glucose 1-dehydrogenase"
/protein_id="AA054105.1"
/db_xref="GI:21449373"
/translation="MPRELAEQVVVVVGASSIGRVTALRFAAGARVVCVCAARNTRAL
DGLVEEVRGAGRAVAVTADIADAEAVPAVADVAVRFGVDTWNAAGIGVYGVVED
TPAGFDRVMVNYLGHVHGAKAALPALURAGGGVLLIGVASVLGRSLAPLQAPYASK

```

Query Match 5.6%; Score 28; DB 1; Length 82746;  
 Best Local Similarity 100.0%; Pred. No. 0.017; Indels 0; Gaps 0;  
 Matches 28; Conservative 0; Mismatches 0;



QY 103 ACCTTCGTGAGTTCTCGCGCAGCGCG 130  
|||||  
Db 22257 ACCTTCGTGAGTTCTCGCGCAGCGCG 22230  
|||||

RESULT 27  
AX153707  
LOCUS AX153707 665 bp DNA linear PAT 22-JUN-2001  
DEFINITION Sequence 31 from Patent WO0140497.  
ACCESSION AX153707  
VERSION AX153707.1 GI:14535358  
KEYWORDS  
SOURCE Streptomyces ambifaciens  
ORGANISM Streptomyces ambifaciens  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
AUTHORS Jeannin, P., Pernodet, J.L., Guerin, M., Simonet, P., Courtois, S.,  
Cappellano, C., Franco, F., Raynal, A., Ball, M., Sezonov, G.,  
Tuphile, K. and Frostegard, A.  
TITLE Method for obtaining nucleic acids from an environment sample,  
resulting nucleic acids and use in synthesis of novel compounds  
JOURNAL Patent: WO 0140497-A 31 07-JUN-2001;  
Aventis Pharma S.A. (FR)  
FEATURES  
source Location/Qualifiers  
1..665  
/organism="Streptomyces ambifaciens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1889"

ORIGIN  
Query Match 5.4%; Score 27; DB 6; Length 665;  
Best Local Similarity 100.0%; Pred. No. 0.082;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 CCGACGCCGCTGCAAGCGTTCGCCG 166  
|||||  
Db 389 CCGACGCCGCTGCAAGCGTTCGCCG 415  
|||||

RESULT 28  
AR362090  
LOCUS AR362090 15872 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 1 from patent US 6600029.  
ACCESSION AR362090  
VERSION AR362090.1 GI:33770246  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 15872)  
AUTHORS Sherman, D.H., Williams, M.D. and Xue, Y.  
TITLE Metabolic engineering of polyhydroxyalkanoate monomer synthases  
JOURNAL Patent: US 6600029-A 1 29-JUL-2003;  
LOCATION/Qualifiers  
source Location/Qualifiers  
1..15872  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 5.4%; Score 27; DB 6; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.06;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTGAGTTCTCGCGCAGCGCG 131  
|||||  
Db 3721 GTTCGTGAGTTCTCGCGCAGCGCG 3747  
|||||

RESULT 29  
AR362091  
LOCUS AR362091 15872 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 3 from patent US 6600029.

ACCESSION AR362091  
VERSION AR362091.1 GI:33770247  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 15872)  
AUTHORS Sherman, D.H., Williams, M.D. and Xue, Y.  
TITLE Metabolic engineering of polyhydroxyalkanoate monomer synthases  
JOURNAL Patent: US 6600029-A 3 29-JUL-2003;  
LOCATION/Qualifiers  
source Location/Qualifiers  
1..15872  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 5.4%; Score 27; DB 6; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.06;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTGAGTTCTCGCGCAGCGCG 131  
|||||  
Db 3721 GTTCGTGAGTTCTCGCGCAGCGCG 3747  
|||||

RESULT 30  
BD217171  
LOCUS BD217171 15872 bp DNA linear PAT 17-JUL-2003  
DEFINITION DNA encoding methymycin and pikromycin.  
ACCESSION BD217171  
VERSION BD217171.1 GI:33026941  
KEYWORDS JP 2002536959-A/1.  
SOURCE Streptomyces venezuelae  
ORGANISM Streptomyces venezuelae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 15872)  
AUTHORS Sherman, D.H., Liu, H.W., Xue, Y. and Zhao, L.  
TITLE DNA encoding methymycin and pikromycin  
JOURNAL Patent: JP 2002536959-A 1 05-NOV-2002;  
REGENTS OF THE UNIVERSITY OF MINNESOTA  
COMMENT OS Streptomyces venezuelae  
PN JP 2002536959-A/1  
PD 05-NOV-2002  
PF 25-JUN-1999 JP 2000557373  
PR 26-JUN-1998 US 09/105537  
PI DAVID H SHERMAN, HUNG WEN LIU, YONGQUAN XUE, LISHAN ZHAO PC  
C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/00, C12P7/ PC  
42, C12P19/62,  
PC C12N15/00, C12N5/00  
CC DNA encoding methymycin and pikromycin  
FH Key Location/Qualifiers  
FT source 1..15872  
FT /organism="Streptomyces venezuelae".  
FEATURES  
source Location/Qualifiers  
1..15872  
/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:54571"

ORIGIN  
Query Match 5.4%; Score 27; DB 6; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.06;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTGAGTTCTCGCGCAGCGCG 131  
|||||  
Db 3721 GTTCGTGAGTTCTCGCGCAGCGCG 3747  
|||||

RESULT 31  
AF082100  
LOCUS AF082100 28732 bp DNA linear BCT 27-OCT-1998

DEFINITION	Streptomyces sp. MA6548 FK506 peptide synthetase (fkbp), FK506 oxidase (fkbo), and FK506 polyketide synthase (fkbb) genes, complete cds.
ACCESSION	AF082100
VERSION	AF082099
KEYWORDS	AF082100.1 GI:3798623
SOURCE	Streptomyces sp. MA6548
ORGANISM	Streptomyces sp. MA6548 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE	1 (bases 1 to 28732) Motamedi,H. and Shaifee,A. The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506 Eur. J. Biochem. 256 (3), 528-534 (1998)
JOURNAL	98451508
MEDLINE	9780228
PUBMED	2 (bases 1 to 28732)
REFERENCE	Motamedi,H. Direct Submission Submitted (04-AUG-1998) Molecular Pharmacology/Immunology, Merck Research Laboratories, P.O.Box 2000 Bldg. RY80M-213, Rahway, NJ 07065, USA
TITLE	
JOURNAL	
FEATURES	Location/Qualifiers 1..28732 /organism="Streptomyces sp. MA6548" /mol_type="genomic DNA" /strain="MA6548" /db_xref="taxon:82632" complement (61..4575) /gene="fkbp" complement (61..4575) /gene="fkbp" /codon_start=1 /transl_table=11 /product="FK506 peptide synthetase" /protein_id="AAC68816.1" /db_xref="GI:3798625" /translation="WTSRGRPGTLRGSGPKRNSNTRITVAADRAEPLLASFAQRRTWFLQQLDPDSNAYNLPLVRITLGPLDAALLERALTIVARHEALRTVFDADGEPVQ RVLPAAPRLIHARADGQGAARLVHDSIAAPPDLSTGPLIRALLIRVGGDDHVLAVT VHHVAADGWSFGLQLELAHCTARLDRSPAVLPPLRYADFAAWERRELTGAGLE KRLAYMRDRGAPALVLDLPTDRPPVADAGLVEWRPPEALAAVRRIRGRDAGTS AFMTLLSAQLVLAHAGTRDVLVGTVPVANRTREFEVIGMVFNTLALRGDLSGPS FRELLDRCRATAFAHADLPENLIEVLAPDRDLSVNVVQVLLVLRDASAPAL PGVTAEPGTGHWFTFRLDEPHVVEADGALTGELLYSRALFDESRIIRLLDEFTAVL RAAITAAGTRLSQLGAEATPTPEIPSDNTARELVPDTLPELLARIYARTPDVAVTDP PCTSLTGELDRRRLAHLKARGTITGDLGVGICADRGVDLVGVGLLKAGAAAYTP LDPEHPERTDFVLADARLTAVVAHETYRHRFSGVPHVTVPDPLGRQPTAPHLTL GRDSLAIYITSGTGKPKAVLMGVSANILLMQERTWGREPASTVQFVTATPDYS VOEIFSALLGGTLVIPPDEARPDPLGARMMDEQAITRIYAPTAVLRALVEHDPSHS OLSARHLCOQGEALADARLELCRHPHLRVHNYGPAESQLITGVTLPADPTWP AAAPGRIDNTRIHLDDALRPVDPGMPGQLCVAGILGARGYLARPELTARWYVSGG ATGERMVTIGLARARPGDDDFLGRIDDQVKIRGIRPEGEVENVLAEDARVAHAA VSVREDPRGEFLAAYVYPTAGPHGGLAASREGLDRLPAYLPVPSAVVMDALPRT SIKGVDRALPDPAAGSPARKVAPRNRARVVCIRFODVLDPQGVQIDDEFALGHH SLIATRVRSIRIAELGDVALRTLFDAPTAALARAADAGAPALPTILSSPCAGPAP ITAAQOMLHSHGSLAASVYTVAPVGFELRGELDRDALDAALTRITARHEPLRTGER GERQIVGPPAAVCAEVSPVSDVGAARVAARAEALRPDDL VAGFLIRLAVLLPLGAED HVLLMLHLDGWSFDLLVRELSPDLSPVDPVAVRWERVFAVVAARNDRAYW RRQFGACAPELPTVRPRTAAPDGRAFWLTDDATVAARARVADARNATWHTVILGA FALVVAEADDTDLVATPFADRGVAGTDLHIGFFAKVIALRIDVSGTGTPSPFAEY LRRVATMTGAHOSVPSYSAARAADPALPTPVSFOLISALSTELRLPLGLTEPPPV VRTVDEINGELSVNLFDDGHTVSGAVVHDAALLDRVDDLLTRVESTLRAAGDLT VRTVGQVGE"
gene	complement (4572..5570) /gene="fkbo"
CDS	complement (4572..5570) /gene="fkbo" /codon_start=1 /transl_table=11 /product="FK506 oxidase" /protein_id="AAC68817.1"

```
/db_xref="GI:3798626"  
/translation="MSAPVAAPCYCRFEKIGSPDLGEGDEVLTGVIEHGTGHTGASLVVG  
VPTALTHTTRDEAEFAEVMHQAOPPVESGTTDDGIAWARTDEYRLFVCGRAVPSGRYADA  
VSYALYTRISGLTSLGHLPLIARTWNVSGINSANADGLEVYRDFCVGRQALDGGID  
PASMPAATGIGTHGGGIACAFIARGGVRIENPAVLTAHHYPAAYGPPSPFARAT  
WLGPPDRGLFVSTAGITAGLHETHHGDVTCQCEVALDNTARVVGAENLRHGGIRRGH  
LLADVHLKVYVRRREDLAAVRRVCAARLSTATVAVLHDTIARGDLLLEGVVR"  
5798..28528  
/gene="fkbb"  
5798..28528  
/gene="fkbb"  
/function="FK506 polyketide backbone synthesis"  
/note="initiation and first four elongation steps"  
/codon_start=1  
/transl_table=11  
/product="FK506 polyketide synthase"  
/protein_id="AAC68815.1"  
/db_xref="GI:3798624"  
/translation="MARAESIRPLHLLRVNAERHGRVAYTDSRRVYTGRLQIRTC  
RIAGHLAASVGRGDRVRLLLGNRIETIEVLAARAAAVPLNPDAGDELAHLFLT  
DSGATVLTDDLQVRRVGTDTATVLAGREFPCVAYEDLACTEPPRPDDLGLD  
EPAMMLYTSGTIGRPGKVSQAORSGILWSATCDVPSWELTEDDELLPAPLHSLGH  
LCILAVLTGASARILGSEVARDVLDALAEHPCTVLGVPTMYRLLAGVSGEPQTRA  
LRVALVAGSTSPASLTAGFEAAFGVPLIDTYGCTETTGSLTANTLTITRVPSCGLPV  
PGLSRFVDPVITGVRKBLIRGENIHPREIESVAQVEPGVTRDAAGAPHPVIGHPVL  
QAETHVITIGRVKBLIRGENIHPREIESVAQVEPGVTRDAAGAPHPVIGHPVL  
YVPEGPVPADALAEBCRLQAYFKVPELITVHTVTPRTASGKVRGELAGLPAFLV  
GTSGEALCELVMERELPDTVSQVAVTRRAVIGPAEVPDQADQLRWDEALVRQ  
AAEPGSLVLDVTDITDNDANDEDELVAALGEPVALRSGVAYVPLVRADIT  
TPLLASQWALRPASGTLRLDAPADPAPPLAAGVEVRVDRAGLNFRLVLTALGT  
YFGDGMGREGAAGIVTEVPGVDLVPAGHVGFLVQDAFRSVVADRLVARIPEGWS  
FVAAASVPIVATAYGLVDAGQVRPGQKDVLIHAATGVGSAQAIAHMGADVYATA  
SPAKHLLADGFDADHANSRTTAFADAPFPVDVNLNSLTGEFLDASVELLAGPGRF  
IEMGKTDIRHDVQPPFDLSVPARLEILLELLELDFDRDELRLPVRPMDRRADA  
FGWMSOARNTGKTVLTPRPLGPDAPVLVTEGADAVAGLRDLETGRTHVSSAAEAE  
RAESEGSGNASTLAHLVLDGPACADAYARSASEGLPIVTVHGLPGTEQAALTCELI  
EAAIEGGSGYVTVRSGAGARAEAMAGATPILSGLAGTGAEPGAGPAGWGSRLAA  
ADARENVLLDVRDVSATVILGLPAGRCAPDRTFRENGLSLITTEFANTVARTGL  
RVPASTDFHTPEPSFAAHLGSAVAAAPAAKALGSEPAIVGMACRLPGGVTSPGL  
WRLVASGTDAITPEFVDRGMDVDELDPDPAVGSRSTTRHGGFTTEATFTDAAEFGIS  
PNEALMDQORLVLETSEWAEFHAGIVPDTLRESDTGVFMGAGHQYGAGRDGLGIG  
VTATQTSVLSRGLSYFFLQGPATVDTCCSSLSVALHQAALRSGCSLALAGGV  
VMAITPGEVFERQSGUSPDGRCKAFADAADGTGFAEGVGVLVVERQDAERNGHTVL  
AIVAGVAVNQDGSNGLSAPGAQQRVIRQALANAGUSGADVDVAEHGHTGVLTGDDP  
TEAQAVALSYQEREVPLLGSVNSVGHQTAAGAGVIRKMWAMRHGLLPKTLHVD  
EPSHVDVAGAEVVTETREPESGCGARRAGVSFGVSNHVLLEGVAESSVRDG  
GSSGIVLPVGSARTESLALOVRIGEYVVGADAAAVADGLVRCRAVDFRVALVLE  
SASVAGAVEGARTVFEFGQSQWVGRELWGASEVFAARMRECAAVLEPHTGWDL  
DVLGEAVVVDVRVLPQASAVAVSALAALWQAHGVVPDAVVGHSOGEIAAACAGALS  
LEDAARVALRSQAIAELRAGRWASIAVPAASAVEVEGVWIAARNGPSTVAGDP  
AAVERVLARYEAGVYRRIADVYASHTPHVEAIEAOLADALEGITSSTFSPWMSV  
DSGWTEPFGAYWYRNLQPVAMDVASELDSGLFIECSAHPVLLPDLQERTVASL  
RTDGDGDRFLAALAAQATQGVADVMTTLEPAPHRVLDLPTYPPDHRYLQLPAPV  
GAGIGHPLSSAVALPGSDGLTKLSLATHPWLADHTVOELPCTGLFLELVVRA  
GDEVGCDTIDELVIEIETHSLPASGAVDITVTVDQDDTGHBPVSHAPAGTDTWTRH  
ATDGLRGTEDAPTNSFPMPPTGAQPSLDEFYERLAGAGTYTGPAQGLRAWRA  
GTVTVAVALDGAADVDFGHPHLLDAAELHAGLDAGDELELFPSTVGRUHLTGAA  
AVRVASHSGPAGVAVQVADLDGRPVSDALVRPSVAPSGPNLLGLEWLEVAEAYH  
DGADLPGYTLITATHPDDDDPTPHNTPTHTQTTVLTALQHLTTNHTLI  
VHTTDPGAAVTGLTRTAQNEHPGRJHLETHHPHTPLTLTQTLTQHPHLPLTNSL  
LHTPHLTPTHTHTTTTTPTPTPLNPHALLITGSGSTGLAGILARHLNHLTYLSR  
TPPTPTPTPTPTCDLTPTQTLTPTQTLTPTQTLTPTQTLTPTQTLTPTQTLTPT  
TQPKADAAAMHHHTQPLTHFVLYSSAAATLGSQGANAYANAFALDALATHRT  
QGPATTTANGWHTTTTTSOLTDTQVRDGRFPLTEAGNEHFLDASLATDAPFV  
VAAIPAEARREVERBERTGTGSDCLLAVVCATAAVALGHADASEITGAATFKDLG  
IDSLGIRLRSLAEATGVRLSATVDFHPTPDALAELELGDSDSQGAAADAAA  
PGCTVAAAGDEPLAIAMACRMPEGVDTPEDLMSLVESGSDATETPTDRGNLAA  
LYDPDPAIDAIKGVSVRHGFLTGAADFDFAEFGISPREALMDPQRLVLEVSWEAFER  
AGILPASVGRSDAGFMGATQYAGVLDLGGFATGTPTSVLSRGLTYVLEGEPSV  
TYVDTASSSLVALHQAARSLRSGCSLALVGGVTVMATTFGVFSRQRLAPDGRAK  
AFADTADCTSPAEGAGVLIVEELSDATRLGHVPLAVRGSVANSQASNGLSAPNGPA  
QQRVIERALDDAGLAPGLDIAVEAGTGTGLDPIEAOALEAYGLDLRHLPLIGLSLK  
SNFGHTQAAAGVAGVAKMWLAWRHGVLPRTLHVDFSPRHVDVGGDVRLLRNEPWPVT
```

GRVRRGVSFGISGAHVVLEAGPAPAPLPAADVPEDVPMVPSARTPDGVRD  
AGRLTALTAAPAAAGHSLATRTAMHRAVVPADAEAFARGEVEVGVVGTADVTDT  
RAYVFPQGSQMDGMAELIATPVPARLIGECALALAPYTGWLDLVIARPGAP  
LDRVDVVPQASFAVMAHAGVAPAAVGHSGVEAGAACVAGVLTIDDAKVV  
ALSRVATERRAGGGMVSPADFDAAWAGRLVAAVNGVAVSVVAGAAVEEILL  
AAYPHARRIVDVASHTAHVSEIRGALLDALDITPGAPBIPFFSTVDEAWLDRPDA  
AATVHARRIVDVASHTAHVSEIRGALLDALDITPGAPBIPFFSTVDEAWLDRPDA  
RGGGARFRSLAELWVRGVSPVSGELRGVPLTPVPRDRYVWDAEPAGTSGHP  
LLGSLVERAGEGAFATLSVRQFWLADHVDGRLIIPGSAVVELLAEGALSGT  
TIAELTVVVDVEEDTHIQTAVTEVSGRRVLRHSRTGAGTWTGATGALSAT  
RDEPFWPPADAEVDTGFTDPLTISYGFAPFAMTAMWAGQRAVASVRLRGQLD  
ARYGLHVLDDAAVHALGTIFADPERRLAFSGSVRIHARAATVDELLERAGDGT  
RIALTDEHGLVLDVGLTVARABEPGTEALFEVAVSVSPVPGWTHLADVPEGDVI  
EGDAGEDVESDPPVVLVAVEPDPSGVRARELGRDLATVTLWLAAPWAS  
RIVVVRIGDPAQALAGLVRKTAGTEHPGRVGRLEADEITPATVEAGLAAGDESHVRV  
TDGTVARLRVTVASTGASPLSGTGLVGTGGLRLVHMLHAGI VDDAVVATVFRQ  
GGRPDGAPDRVRYVADVTRDELAALVDGVAERLCVAVHAGI VDDAVVATVFRQ  
QMDAVLRKADVAMQWHLTRDLAALVFLVSSISATFGGAGQANYATGNALFDALAR  
HRDQGLPAYSLAWGLDEADMGRLAATDTRIARGGMIPTMAQGLELFDALHS  
DRAALVRLDLAAVASDQVPIRLDLPVNRARAAATPAARDMLELVRTSAALVIGH  
RDHAREPTAFKEVGFEDSLGVELNRNADATGTLTPATVFDHPQAALAAHDEL  
AGATATRTTAPARHDEPLAIVMACRLPGGVASPDLLWRLLSGDGTITAPAD  
RGWDVADLYDPPDPSHSTVVRHGGFTGAGDFAGFGISPREALANDPQVRVLET  
SWEALEHAGIDPHTLGTGTGTFVGMGYFYGSGADRGDFGATSTQTSVLSGRISFY  
CLEGPATVTDACSSVALHAGOSLRTGCSIALVGGVTVMASPSGVDFSOQGL  
APDRCKAPAAAGDTAFAGSGLVVERLSDAERHGRVLAIVRGSVFNQDQANGL  
SAPNGSOVRVIRGALANAGLQPSDVAIEAHGTGRLGDPTEATALLATYQDRATP  
LLGLSKSNGHTQAAGVAGI IKNVLMWRHGTUPTTLUHVDTPPSHVDSAGTVELLT  
DARTWETGRHRAVSSVSGTNAHV IESHPRTTAPDPTGSSVTHVPLLSARTP  
EALDEHTRVTRFADGDERAVASALTTRTAFTHRAALIGTDLITGTAEPDRRLW  
LFSGOSRPGMGDLAAAYDVARTRDVLDALQVAGLDVHDTGYAQPVAFAVQVA  
LSAQDDAGVRDVLVGHISIGELAAVAVGWSLDDATELSARALMQLPQGMAMA  
AVSASERDALPLCGEVAEVAAGVSPVLSGDEADVLDVAARLGRFTLELTSFAHS  
ARMEMPLEDFRDVAERLTYPKPMAGADCTPEYVVRQVDTVREABQVAYDGA  
ALLETGPRLNARLVGIPVLHGEDEARSAMTALRLHGGVAVDVEVIGAPTDLP  
HLTPYPRTRVWLSRAAGDAAPAGQLPVAPVLTAAVMVPGTGLVLTGRVEATDP  
LATRVHPCGAARVLLDRSRATKAAPWRSPWTPRSALPQSGGLALSVTVAAPG  
EDGRVANVHARPGTGEWTHATGILAVARVASVAPETPSWPPATARPFDVGBLAD  
RIARAGHTGDPALPRAANADDAAVHAEVALADHQAADAERYGLHPALLGAALAG  
EGADLPFADVYVHATGATTVAVTATGTHLADTSGVAVAVRNRPLTVEGAV  
PGLLPDLIAELPPTTATGGLDDVPVVDVILPAHGTGGGPGTGERDRLGVAA  
LRSFLTDREYADAVLAHVTPGLAPAAAGVRLTAQAEHPRIVVVDAEPDTPAELLA  
AAGLGEPOVTVREGRVARRLTTPAVPSGDAPELDPDGLVITGSGTLAGIVARHIV  
GHYVRLMLLGRGTSADVPGAEVAVACDIADRELASAGIDPAHPLFANVHTA  
AVIDDPVLTALPRLDVLPRKADGAWHLHEITKQDLAALCSTFPEPCSTERDA  
TTPRRTLLDEAEORRAGGLPALSVANGWEPSGLTAGAGDRMRDGVLTALSTERGL  
ALDLAALRSPALVAADPGGLGSSALLAPHRPRRAAGSALADRIAGISAADAE  
KAALAVRCAAVLHGEHPARIEATATFELGVDSLTALVRLNATFEATDVLPAATA  
VDFPPTQAVAKLTALTCVRPARATKAVGHDPELAI VGMACRLPGVASPDLL  
WRLLSGGGGICITAFADRGWDVADLYDDBEHSGTSTVRHGGFTGAGDFDAGFGIS  
PREALAMDPOQRVLETSWEALHQAQIVFTGLRGSTGVFMGAFSDGYGLGTDLGGG  
ATGTQTSVLSGRISYFLEGGPVTVDITACSSVLAHQAGSLRTGBCSSALVGGVT  
VNASPGGVFEFSQQRGLAPDRCKAFEAADAGTAFAGSGVLTVDASERHGRVLI

Query Match 5.4%; Score 27; DB 1; Length 28732;  
Best Local Similarity 100.0%; Pred. No. 0.057;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 280 GTCAACTCCGACGGCGGTGCAACGGC 306  
|||||  
Db 14927 GTCAACTCCGACGGCGGTGCAACGGC 14953  
|||||  
RESULT 32  
SPAS580915 74787 bp DNA linear BCT 10-DEC-2003  
LOCUS Streptomyces parvulus Tu4055 cluster for borrelidin biosynthesis.  
DEFINITION  
ACCESSION AJ580915  
VERSION AJ580915.1 GI:39725416  
KEYWORDS 2-hydroxyhept-2,4-diene-1,7-dioate isomerase; actolactate synthase;  
aminotransferase; borA1 gene; borA2 gene; borA3 gene; borA4 gene;  
borA5 gene; borA6 gene; borB gene; borC gene; borD gene; borE gene;  
borF gene; borG gene; borH gene; borI gene; borJ gene; borK gene;

borL gene; borM gene; borN gene; borO gene; borrelidin polyketide synthase, type I; cytochrome p450; dehydratase; dehydrogenase; ketoreductase; muconate cyclo-isomerase/lactamase; ORFB3; oxidoreductase; thioesterase type II; Threonyl-tRNA synthetase. Streptomyces parvulus  
Streptomyces parvulus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
1  
Olano, C., Wilkinson, B., Moss, S., Brana, A.F., Mendez, C., Leadlay, P.F. and Salas, J.A.  
Evidence from engineered gene fusions for the repeated use of a module in a modular polyketide synthase  
Chem. Commun. 22, 2780-2782 (2003)  
2  
Olano, C., Wilkinson, B., Sanchez, C., Moss, S., Sheridan, R., Math, V., Weston, A.J., Brana, A.F., Martin, C.J., Oliynyk, M., Mendez, C., Leadlay, P.F. and Salas, J.A.  
Biosynthesis of the angioinogenesis inhibitor borrelidin by Streptomyces parvulus s Tu4055: cluster analysis and assignment of functions  
Chem. Biol.  
3 (bases 1 to 74787)  
Olano, C.  
Direct Submission  
Submitted (19-AUG-2003) Olano C., Biologia Funcional and IUOPA, Universidad de Oviedo, Julian Claveria s/n, 33006 Oviedo, SPAIN  
Location/Qualifiers  
1. 74787  
/organism="Streptomyces parvulus"  
/mol\_type="genomic DNA"  
/strain="Tu4055"  
/db\_xref="taxon:146923"  
/note="Streptomyces parvulus"  
2. 313  
/note="ORFB1"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAE45653.1"  
/db\_xref="GI:39725417"  
/translation="IPRGIVDVRAGRWPHSAAPRHRGLLNANWGANWVATVDFRY ASRTYDDQDVDAIHDAAGLVNAGAGFDILAVALLFFVRLTAQAHAAGALPTPT H"  
501. 3107  
/note="ORFB2"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAE45654.1"  
/db\_xref="GI:39725418"  
/translation="MEAFLLAAESVLLRRDQSVVYTPGSEDPGPAPRAALRUEALELL GRHVASPLHVAQLSEELAAHVRVLVGLVDDLLGSDRTHTPLFRFPRTVFRDT EALYDVRVAFLLQQEQPCVLCGEARTVLPVSPCAHLVCLRCWDSGDYAGCPLCHRR RLIALPLTPAGGDLPPQIPVRETRKALVGLVRRAPSPAPRLLELLELTTATDVL RLIALVSGGDLVTPARTNFRSLRDLAVLDLPAPYLVEDMLRHPTANKRAE WHPFFGHTVRHRAALATAVIRATPLDPTDPTAFGALLTTAAHPDAVVRPDTVRKPT VAGRLQMAEGDAARAALAGERPGLVRLDLVLRHTDEALVPELEKALRHGLPK VGGPLLALGALRTEDTGTRVFFPRGDDVTRALSVPERRPALPAGVSEVALL EGELRRFAAGRPDLSVDAFFDAEWTTGLCDTNLVHGDAAIHSGDLTSAAPRGA FLWTEFRGNRTDLSVAFDDAEWTTGLCDTNLVHGDAAIHSGDLTSAAPRGA TRVYFDLGRSKVCLPMI VDLARRALWTDTHLPSAGFQSI GSHGGELAAVAGDLM QQTSGRATLWDLAVLRAALSPVAVVSRPEPAVLYRRAAESEAFVAVASH KDAERLAHTDPSAAGLAAGRRVFLATVHGDVVRPGASGTSYRLFPAGDASPTILT RVTAGDILLALG"  
complement (3172. 3810)  
/note="ORFB3"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAE45655.1"

```

/db_xref="GI:39725419"
/translation="MAEQIAGIEIPDSAPREATDLIRDTTPPLIFHRRVVLFGSL
QAAALGIRPDPELLYIAALFHDITGLVPPYRGDQREMDQAQAHAFLLAIGIPRADA
DTWTAVALHTTEFVYRPAPEIATAGTETDVLGLRGNLITRAQIDAVTAHPRPD
FKQILRAHTEGFEHRPATTFGTGNADVLEHFAFGERRTDFEVIENSAPPE"
complement (3935..4924)
/notes="ORFB4"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE45656.1"
/db_xref="GI:39725420"
/translation="MTARASHGILVFDGMKMLDLSGPAEVAFAENRFGARLYRVGS
PDGAPVRSIGLIVPAEADAGAPDPTLVVGGDALPGSPDPRILIDAAKALAAAPG
RVASVCTGAFVLAAGLEGRATTHOHTALARECPSTRVEPDALFVKDQATVISA
GVTAGDIDALALLEEDHGPLARVARSUVYLORAGGQSASLRGAPRTPVLRLQ
VQDAQADPAADHSLAALAAARVRSRPHLTRMFAELDVTVPKYVELIRPDLAKALID
SGHNATEAALSGFPSEYSLRRAPARHLGLSLPTRYRQRFATTPVDAGPRPDG"
5123..5953
/notes="ORFB5"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE45657.1"
/db_xref="GI:39725421"
/translation="MGTVTTSDGTSTFYKDWGPRDAPPIVFHHGWPLTADDNDOMLF
ELSHGVRVTAHDERGHRGSGQSTGHMDTYAADVAALTEALDLRDVAHIGHSTGGGE
VARYARAEFGRVAKAVLVGAVPPVMQKSDANPGPTPIEVDFDGFRTALANRAQFYID
VPSGPFYRGKAVSQGLIDNMVQKSGAANAHYECIKAFSEFTDFTDLKADVP
VLVAHGTDDQVVPYDASPLVSFKLLKNGTLKSYEGLPHGMLSTHPEVAVNPDLDFVRS
"
complement (5961..6518)
/notes="ORFB6"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE45658.1"
/db_xref="GI:39725422"
/translation="MGVMTGPAGRERDEGHVTVQAPVATDERRVVDKOTPGCAKAF
VAAASVREMAAAGLDRLVELNVRVSQLNAACVLSHTAAALRAGETORLVAL
PAWRDTEFTAREERAAALAEATRTPADAAQAQAAQARGVLSDDSEVSATVIAISI
NAFNRVSLSKHPVRGAAPAPVSPA"
6567..7538
/notes="ORFB7"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE45659.1"
/db_xref="GI:39725423"
/translation="MSNTEPRPAEMRCGALEDEVPAGVVELVTARDVPLGPRAMTVR
RTLQPARLTIGAWCPCADHYGPDVVAASGMDVAPHPHIGLOTVSMVLSGFEIHRDLSL
GTHAFVRPGLNLMVMTGGFIAHSEVSTPDTVLHGVLQVVALPEERHDTGRDQHPAP
APVAFDGTARVPLGSLAGTSPVSTFTPLGAEALLVPGGTATLDVDPGFEGVLVD
SGDVRVEGAVRPAELGYAPAGRTLTLTINESAAPRLILGPGPPPEEIMWNFTIG
RSHDEIVRAREDMKGRFGEVHGDCAPLPAPELNPAPLKPRRRAR"
complement (7603..8397)
/genes="borB"
complement (7603..8397)
/genes="borB"
/function="borrelidin biosynthesis"
/codon_start=1
/transl_table=11
/product="thioesterase type II"
/protein_id="CAE45660.1"
/db_xref="GI:39725424"
/translation="MTGNTNTHSDVWIRQYRPAHTAPQLICLPHAGGSAFTYHPVAA
LAPCDVLAVQYQORORAEKPLEIDELANQLFPVLRARVHOPVALFGHSGMGTALA
FELARVESAGISEALLVSRAPAPSPQRTGTVHLLSDRELVAELTLDGTAEOVPH
DEELVRMALPAIRDYRAATRYRPPGKRCPIHALTGDDDDPMTVPVPEARWSEHTD
GPFTLDTFAGHGFYLLEHREDAILGIIAHLRTGSRAPGDRSGULTRE"
complement (8397..9194)
/genes="borC"
complement (8397..9194)

```

```

/genes="borC"
/function="borrelidin biosynthesis"
/codon_start=1
/transl_table=11
/product="putative ketoreductase"
/protein_id="CAE45661.1"
/db_xref="GI:39725425"
/translation="MSLELTDVVMVTGAGSIGIRAAARLLVGHGARVVLVGRTESAL
TETTAGLPSHHLVVPCDVDDKQVADCVARAVSRFRLDDGAFNNAGTFTGSGPLHQD
TADNFDRTATNLRGVMSCHRGQIEAMLTAGGAIWNCASVAGHAGOSPLYSAFKH
EVLGSLKVALQYAGDGIKRVNVSPGSTPTPMRLSILYADPSALAQARAPRLGRKGC
EVANAVVLLSPLAAYVTGTGVDGVTAGSAIPRITNATPEGQHR"
complement (9244..9996)
/genes="borD"
complement (9244..9996)
/function="borrelidin biosynthesis"
/codon_start=1
/transl_table=11
/product="putative ketoreductase"
/protein_id="CAE45662.1"
/db_xref="GI:39725426"
/translation="MTARHDLVALVTCAGSGICAEVARGIAARGLRVLLDKDAEAVHR
VADGLDLRLARDPLVADVTDPHALASAVDSLAPQHRPGLVNGVGGDTARSTVLEITE
ADLQEAFTNLASVFTMTRLCVPAMVAGMRVYNLASVAGRTYTRFSNAAVVAAGAG
VIGFTKQCAVELAPHGVTNVAVHAGVIGTERIRRAWEDKPPQWTDADVSHIPAGRFGS
VAEAGMVCHGAGEDAGYTTGTVDVNGGLHI"
complement (9993..11165)
/genes="borE"
complement (9993..11165)
/function="borrelidin biosynthesis"
/codon_start=1
/transl_table=11
/product="putative muconate cyclo-isomerase/lactamase"
/protein_id="CAE45663.1"
/db_xref="GI:39725427"
/translation="MIRVRLHTAVVPMAAAFDHAIRSRRAASLLVEIELAGTRMGW
EGAPRVDVGETLDGAVRAVOACDPCGELAEIEWRDFESAVASIAQLPLTGLVDGSSA
AAVEALLDLADNCRHEAPLADVLRLVAPPASRRDGPSTSVLHLSRDVATVLDAL
TPFALAALRHVKIKVADPAGAVDRLTAQDRLPADTRVLDVNGANTAEAEKAVGEL
DGVGVEBPLPPRSPWELGKRRATLGVMLDESCTGPADLHAAATSGMAASHINVRLS
KCGGFLAAALRLARDELGVGCOLGVHVAEVLGPAWAGETLATAMDLMQTVAGRADE
WFEVPLTTPATVDRSLHVRVEPLTGPCTGTEELLRTRCAATWESGGGMRRT"
complement (11162..11980)
/genes="borF"

```

Query Match 5.4%; Score 27; DB 1; Length 74787;  
Best Local Similarity 100.0%; Pred. No. 0.052;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	136	GCCCCGAGCGCGCGTCGACGGCGTTC	162
Db	34938	GCCCCGAGCGCGCGTCGACGGCGTTC	34964

RESULT 33  
AB070940  
LOCUS AB070940 104326 bp DNA linear BCT 22-JAN-2002  
DEFINITION Streptomyces avermitilis oligomycin biosynthetic gene cluster.  
ACCESSION AB070940  
VERSION AB070940.1 GI:15023967  
KEYWORDS Streptomyces avermitilis  
SOURCE Streptomyces avermitilis  
ORGANISM Streptomyces avermitilis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,  
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,  
Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.  
TITLE Genome sequence of an industrial microorganism Streptomyces  
avermitilis: deducing the ability of producing secondary



/translation="MACRLPFGAAGPAAEFLLALLRNGEDAVCAAPGDRRYAPRRGGFLDS  
VDFEAGFGFISREAAADPOORLVLELSWEALDAGIPPARLGGSGAGVFGVAD  
DYATLSRAAGADAATPQTGTGNGRIIANRVSYTLGLRGPSTFVDSGSSSLVAHVLA  
ASLSRGESVALAGWNILIAEDSVAVRFGALS PDGYCYTFDRANGYVRGEGGG  
VVVLKPLADAIADDHICILAGSANNVGGEGGLTPDRAGOEAVLITAEQAGLSP  
AAVGVYELGTGTPTGDPVEAAAVAGVAGPAGQPLVGVKYNIGLEGAAGTAG  
LKAVLAVRHRIIPATLTHASPNPAIPMAELGTGNTALTTPWEQAGPLVGVSSFGM  
GGTNCVLTPEAGYDAAAPATAGIPAGTAPILVSGRDEQALRAQKRLDHDT  
HGPLRDLVAHTLAARAPLAREAVLLAADREDVSLAGRLADGTGADLVKRTVSPG  
ALAFVTGQSGRQPCMTABLYRTSAEYAAALDEVCSLDPLWRRLRDLVFAEGSDE  
AALLDRTFOPALFVAVLFRYAHCHGLTFRFLGHVSVELAAHVAAGVLSLADAC  
LVAAARGMLQDRPATGAMVSQVATDELAPFVSESIVLAAALNGPASTVIAGDEDV  
AVAAHWAARGKTKRLRVSHAFSPHMDGMLTFERRIAAGLTFGAPRIITVSNETGL  
LTDADARSPEYVHRAARVFLDGLVRLLEDQCVTTLLLELPGDGLVSAMARDCIRAG  
TVCPIILRGSRDPTVEPALLAAGLVYRVPMDERTATAGARRYPVLPYAFORRRH  
WLPEPASASRPVNGGQAGHAGQAGQAGQAGQAGHAGHAGHAGHAGHAGHAGHAGH  
DRTTLESDRATVQVIGHASPDVVAHEAPRELFGSLMIAELGERLITATGRVPTT  
LUDHPTFDALARRLEADGAERTDAGFVSPAAAHHDDPIVIVGMGCRPLPGVRSPE  
DLWRFVAAAGDAI SALPVDGRMEFAADFFEGGFLADVAEFADGFGFISPREALMDP  
QORLLLETWEALERAGVDALKLRSGRTGVFGASPSYGRHETSDADGHVLTGTA  
PSVLGRIAVVLEGPALTVDACSSSLVALHAAQALRGEGCDLALAGVTVMATIS  
GMPAEPARQGLAPDRCRKA PADAGDGTGKGGVGMVLAQRLSDAVRDCRPLVAVRG  
SAVNSDASNGLTAPNGSQORVIRQALASAGLSAADVMVEAHGRTGSLCDPIEAQA  
LLATYGODRPPDRPLMGSNIGHTQAAAGVGVKSVLAUNGLLPKTLHDEPT  
PEYDWSAGAVELLTEGREWPEPTDGPRRAGVSAGFISGTNAHVILEQAGSYEEVETAPA  
DDGLLVFWVVSARSEALRAQARRLADHVRESGLPADVGLSLAVTAGLEHRAVLVA  
EDCAGFLAHUADAEGRTAGVVEGTGTAFLPGQSGQRMGRELVSAPHPVFAAFD  
EVCAPFDGLLPGLAREVVEGGAGLDREWAQPLALEVALFELVSMGVRADVLIGH  
SIGELAAVYVAGVNSLADACRIVAAGRLMQALPEGGMVAVAAADELDPDLPGVSV  
AAVNGPRSLVSGDESEPTALQAFQAGRTKQLTIVSHAFSARMEFPLAEFAETIA  
AVFRPFRIPVNSGVNAGADEFTTPAYVVRHVREARFADGVGAHAAAGVRYLEL  
GPDGTLTAAHTADAGROVAAPALRGDQPTRTVTVAVGLYATGLTVDWPAFFA  
PYGARPTLTPAFQHRVLEVPVAPSSAPANPLRYRTAWASVPDGPRLSGTAWA  
VVPASGYEDIVAGVTAALRDHGAPVWVVDVDPAGPERADVATLRAALGAGGVVS  
LLTGAGLSGQATVATVQAVTDLGGARLWCLTRGASVSPSDPLAAAPQAQMGVM  
GRVALEHPRGWLGLVDPVDDPDRATRLCAVLGSGSDQVAVRAAGLFAARLIGA  
GTADAPAPGSPWRDTGTVLVTGGTGAHLAQMLAAAGAAHVLITSRGSPDADGA  
AELTARLEGTGVTVAACTVDRLDALAELLGLPADRPLTGVVHAAGVLDGGLVDSL  
TPDRFAAVAGPKVTGAHHLHELTDLSLAFVLFSSPAGTVGLAGQANYAAANAYLDA

Query Match 5.4%; Score 27; DB 1; Length 104326;  
Best Local Similarity 100.0%; Pred. No. 0.05;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 CCGACGCGCGTGCAGCGTTCGCCG 166  
|||||  
Db 34965 CCGACGCGCGTGCAGCGTTCGCCG 34991

RESULT 34  
AP005033/c  
LOCUS AP005033 320150 bp DNA linear BCT 10-MAY-2003  
DEFINITION Streptomyces avermitilis genomic DNA, complete genome, section 13/30.  
ACCESSION AP005033 BA000030  
VERSION AP005033.1 GI:29606546  
KEYWORDS Streptomyces avermitilis MA-4680  
SOURCE Streptomyces avermitilis MA-4680  
ORGANISM Streptomyces avermitilis MA-4680  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M. Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
2 21477403  
PUBMED 11572948  
JOURNAL  
MEDLINE  
PUBMED  
AUTHORS Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H.,

Shiba, T., Sakaki, Y., Hattori, M. and Omura, S. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis Nat. Biotechnol. 21 (5), 526-531 (2003)  
22698306  
12692562  
3 (bases 1 to 320150)  
Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, N., Director-General of Biotechnology Center, Shiba, T., Sakaki, Y. and Hattori, M.  
Direct Submission

Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/).

Tel: 81-3-3481-1933, Fax: 81-3-3481-8424

This work was done in collaboration with Haruo Ikeda (\*1), Jun Ishikawa (\*2), Akiharu Hanamoto (\*3), Chigusa Takahashi (\*3), Mayumi Shinose (\*3), Hiroshi Horikawa (\*4), Hidekazu Nakazawa (\*4), Tomomi Osonoe (\*4), Norihiro Kishida (\*4), Hisashi Kikuchi (\*4), Tadayoshi Shiba (\*5), Yoshiyuki Sakaki (\*6, \*7), Masahira Hattori (\*1, \*7) and Satoshi Omura (\*1, \*3).  
Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.

- \*1 Kitasato Institute for Life Sciences, Kitasato University
- \*2 National Institute of Infectious Diseases
- \*3 The Kitasato Institute
- \*4 National Institute of Technology and Evaluation
- \*5 School of Science, Kitasato University
- \*6 Institute of Medical Science, University of Tokyo
- \*7 RIKEN, Genomic Sciences Center

Following url is also available.

http://avermitilis.ls.kitasato-u.ac.jp.

FEATURES  
source

1. 320150  
/organism="Streptomyces avermitilis MA-4680"  
/mol\_type="genomic DNA"  
/strain="MA-4680"  
/db\_xref="taxon:227882"  
/note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRU 8165.--synonym: Streptomyces avermectinus"  
complement (134. .11953)  
/gene="olmA3"  
complement (134. .11953)  
/gene="olmA3"  
/note="SAV2897 oligomycin gene cluster"  
/codon\_start=1  
/transl\_table=11  
/product="modular polyketide synthase"  
/protein\_id="BAC70608.1"  
/db\_xref="GI:29606547"  
/translation="MANEQQLDYLRKVSADLHRTORLTETIESQGTETIAVGMACH YPGVSSPDELWRLVTDTEIGFPTDRGMDLGDHPPDPHFGTCTYABAGFVEDI ASFDAAFGISPREAQMDPQRLLETWEALHAGLDIHALRSGTGVFAGLNQOD YGTLLAAAPDGLDSYGTSTNSVLGSLGEPVAVTVDPACSSSLVTLHAAQ ALPSGCDLALAGAGTTLTSAVHLSGQALAPDGRSKAFSAADGAGWSEGVGL AVERLSADRLGRHVLAIVLGSANQDASNGLTAPNGSQORVIRQALANAGVSPAE VDVVEAHGRTGSDPFIADALLATYGOAKTAGRLPWIGLSKNSIGHSGAAVSGVIG KMVQALRGHLLPTLTHAEPTPNVDSGSAVELLTPARPASDTPRLAAVSSFGIS LNAHVLEQAPETPEAPEETPEAGATAPLLPWLVGARGEPALRAQARALREHL LARPGSADVALSLAVRTGLREHRAVLAARDGFLDADLATDRPAEGLVLRGSP LVDKVVFEVPGQSQWAGARELLDTPVFAAKAAACAAAEIEFVDMKNKLDVLRKPG AASMERIEVQPLFTVMVSLAELWRSYGVPEPDAVSSSQEIAAHVAGSLTQDAA DALAEFTAEACARDNVRARIVAATVPTCALVDPLRLERLLDLAPVARTGVPYISTV TGGRLDTAAMDAGVYNDTRPVLFPVVRVALLAEHSAFVSSAHGAVLQVETVD AAGASGVVSELRDEGLGRMTASLAKHLGGVVRVDVPAFAGTGAETVELPTVAFQ RTRVWTAGRTDVGSGVLSVDHPLLGLVLRMADGDGAVLTCRLSIRACTWLAGHG VSDRVIFPGTGVFELAVLAGQVCGGRIBELTLHTPLVPSGAVTVQAQVEADETG ARAVSVYSRADDSADTVWTRHAGVLAEPGEPGEGITGTWPPEGAEAVDISGMYDR





```
385..13257
/genes="myca ORF1"
/notes="putative"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="BAA76543.1"
/db_xref="GI:4586928"
/translations="MSRLPQASGVBEFMDLLTEGRDAVARQDQGSWRAALAEFTGFD
AGFGMSPROAAADPOORATVLELGEALHAGVVPATLAGSRTGVFVGVYADYAT
LWRTCTTPDGYTTPGLNRSMIANRLSYLLELRGSPFWVDSAQSSLVVGHVLAACESLR
GESDAIVAGVSTLSTPSTAAWASGALSFDGCTTFDARANGYVRGEGGAANVLR
LSDALVDGRKHVAILGGAVNDDGGPGLIAPRAAQEEVIRLRDRSGVGAQVTLV
ELRGTTGVPDGPVFAAALGAGADRRADPLHVSXKNSIGHLEAGAGTIGLXKAVL
CVREGLPPTLNTQTPNGIPLAQLRLVQTVTRWPRDRADEGRIRAAVSFGMGGTIN
AHLVLAQAPRAVPAVSPTTELPAVFWLVSARSEALRQAEERLREHLRDLPLRPL
DVGFSLATSDVPHRSVRPADGEPVAAAGSRLAEGVAVFPQCGAQTGMAAASLT
ESPFAARLTCACAAVTPFVDSLLDVLRAVPGSPRLDRDDVVQPAFCFANVSLAEIL
RSGWVPEAVVGHSGEIAAANVAGALSLEDGARIVALRSQLLTREHLGGWVSGVL
PVAETELRLRWGRHLAVVNGFNATVAVAGDGLDELTECERQARVRVAVEYA
SHSPHYERIRDELMTALRPVTPQSAFYSTADVGTIDGERLDGAYRNLRQPVRF
ADLRQMSDTGYTSFVELSAHPMLVTLAEILAQETTAALCVTGSRLRDEGLRMLA
SAAEGWAGLIPDRWKEVAGCTGAEVLEPTVAQORYWAPDAERTAVPAAVDAPVR
PAASPEQATPAQATQQTIDLLKLTLELTVAIVGQSPSAETVDPDLTFKQGLESTVA
VFPNRRLARVGLSLPAILVDFPTTAVAGFLGDLHPRESQLVTVPFALVDDDEPIA
IVMGCRFPFGVSPEDLMDVAVASVDVAFPPVDRGMDLESFDDGPDRAAGTYCRB
GGFLDDAAGFDAGDFGVSPREAVAMDPOQLLLEGGVVGAGGFSAGVGRSGRTGVY
AGVMOBYGRLYESABEHRGYLLTGTAASVSGRVSIVLGLGPAVTVDTAGSSILV
AHLVLAQALSGEDNALAGVTVMAFGI FVSESROGLAVDGRCKAYSDSADGTCM
SEGUVVVVQVDRSDANAGRRVLA VRGSVINQDQASNGLITAPSGSQQRVIRAALAS
ARLSVDVVEHGTGTRLDPIEAQALLATYIRGGGRSLLILGSKNSIRKIAQAAG
VAGVVKMVLAMERGVSPASHLVRCAVACQDSSGAVVEVGEACAMETGCRPRRAVS
SFGVGTNNAHVLLEQAPVVDVADSSAGAPVASVSDADSSVSPVWLSARSWGLRA
SAGRLAALGPIGLSGSGVSGVSDVGLSGRSVFGWGRVLSGVLGVLGVLGG
VGSVGVSGRSVAGGVVFPVPCOGAOMVGWGLCSGVSEAEDEVGVQGLVD
WPRGTACREVRGDLGRVDVQPLSFVMVALARLSRSGVVPVPGWVGHSGEIAAA
VSGSVALVEGARGVGRSVI GARLGGAMVWVGASVGVGEGLLGGLGGLGAVV
NPGQVGVAGDGGGGVAGGVGNGVRRARLLPVGSGHGLVDVVRDELLGVLG
RGSSVTFPFFSVGVGVGVDGVDGYYFNRLRVRFGVVRVLRVGDGSSFSFEVSA
HPLVGAIEVDLGVGGVGGVFCGFFVAGVGVGGGLGRVGVSAAGVGRGLPVWRVVP
LAGGRCVGLPTVPHOHDVYLAPEPTAAVAGATAADGAAAGVGRVYRWGKALVPV
TPGRLGRNVLVHGHTGVGDLQSGVADILTRHGATTEHLTIDPAPSGVLAERLGL
AAAGEPVRVSDLLDGLLVQAMGRAGLDARIRAVTVGAVASDEVSPESSEAL
WAFGRVALELIPORWAGLIDLPVQADARALERLAGALTSEENQALVRAAGLYTRLI
RTAGRAKEWRPRTVLVTGTGALGARVARMLADHGAELVLVSRGCAATGAELV
AGLERGVTVTVAACDVADRDGACRPAGGATAGLPHRWRRAPRDCGPHASRGGG
VPAERLAATHLHDLDIDAFLVSSIVGSMGNAGQAAVAAANAALDALAEORRGA
GLPATSLANGLMAGADGAGEQTFRGVAMDPPDDGTALALQALDSGTCVTVAD
VWPSMVRTHANPAARLFEIPEARQALADRAAEQSAERPSDLARLTLGUTADRRL
LLDLIGAQAIVLRQDSGWTIPPSAFRSGFDSLTALERLNRNATGLTLPSTVV
FDHSPVALADFLDVLPRSGGADVEVAAAPDEPIAVIGVCYRPGVGVAPTEQLW
DLILTERTDGLPTDRGMNDIDTDPPEGRVGRTYCREGFLHDADFOAAPFGISF
REALMDQORLLLETSEALERAGIDPRSLGRSRTGYTCMTHQEVAAHLHEMPEY
EGHLLTSGSASGRSVYVLGLGPVAVTACSSILVALHLVAQALRGECDLALA
GQVTVMATDCSWSPASGASPRPVQVLRQADGTGSEGAVITLARLSDVDRG
HDTEAGALLATYQGRFPVQVRLGTLKSNIGHTQAAAGVAGMKMVALRHGVLP
TLHATPSHPDNAGVSLVLTERTDWPDPNPRPRAAVSAFGISGTNAHILLEOAPF
TVPSERSEARAGLVVPLSGRSEPALRAQARLDFLDGADPDVDFGSFATTR
SRLDHRVLLDGTATEALHALVSGDAGRAVLSGTPDTALFLFTGQSSQRLGMA
DGIRPEFPFAEALAEIAEMDRLDQPLLSVMFAEPGSATATILDRTEYAQPALFV
EVALFRFESWGVRDVLVLSHGEIAAHVAGVMDLEHACRLVAARGLRMQMPADG
AMIAIRIPEERVRQLMDLGRHVSVAANGPESVVSAGHVSRSLADQLTAEGHRVR
RLNYSAPHSPHVDVLEEFKRAAAEVYRGLPVVSVNVTGAEVYRATEDLSAEYVVR
HLRSPVPHDGI RLLDRGATLVELGPDGVLGCLAAQCLDPRDEVALPALHRTSG
EVRSLNAAATVVRCHVDVNGGFETGRSLSLAAQPLVIAFORQRYWASVSPGSDQTA
AVRFLGNADHPLGGSLASVAGSDILFTGRSLSLAAQPLVIAFORQRYWASVSPGSDQTA
ASRAGVAGGVRDRLDQLEAPLVVPRDRAQVQVVLTCGDEPRDGRQVRVFAPEGSTA
PWTRHATGNGAAGSGRRPGVGVALPEGAVALDGLVDRYRFAAQHLEYDGVGTGVR
ELWRADVFAEVALPDGAEEAAFGVHPALLDAALQPLVAGVDRPADSPSLRPLFA
WQGMQVTSGAALVRIRRRAGESAMQVAVDGA PVCSDALVLRPTTVRSRTTV
GPALLPLHLRWREREPQAGTFLTFLPTGAEERYGAEVTVHQDLTALRAVQORGA
PPPRAVLAAGFSGPHRTAEAVRAVAAAAALDLVQRWLADDRHQDSRLVLVTERAVAADH
```

EVNPDLAGAAVWGLVRSASEHPGFALIDHDTQESRALLPAVALGCAQLARQGT  
VRVPAALPPTESGADTRAPDGTIVTCAAGALGALVVRHLVAGHVRLLLSRH  
GREAGWABLAEDLAQRGVTVTIAQDVADRTGLAAALDAVDPDEHPLTAVHRAGVLD  
DAMVTCGLTADRLDEVLRPKVDAALHLHETADRKLAAFLVFAAAGLLNPGQASVAA  
ANTVLDALAHBRANGLPATSLAWGLVGGPENGMAANLAHLRLARYGTIAPLDGDEA  
RSLLDALRTDRALLVPLRLTRVDARQDPADVEIILRPLVFPVADGGGSPATGNS  
LEAFRAQALAEDDGTGEHLLDLVRRQVAVELDRADVDDPKRFPQDLGFDLSLLAVE  
LNRRLQATGCLPLPATIGFDPHTPTQVPHLAESLLSRPEGDPGEQALGLDAVARAL  
TKLGDIDIRVVVVRRRLAEALLTAVAGAEPLDANGGTGAGVDGDHDDTALVERLSGAS  
DELLAPIDEQI"

ORIGIN

Query Match 5.2%; Score 26; DB 1; Length 17101;  
Best Local Similarity 100.0%; Pred. No. 0.18;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 385 GTCCAGCGCGTGGAGGCGCACGCCAC 410  
DB 8656 GTCCAGCGCGTGGAGGCGCACGCCAC 8681

RESULT 36  
AF016585 41097 bp DNA linear BCT 07-DEC-1997  
LOCUS Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nidi)  
DEFINITION gene, partial cds; polyketide synthase modules 1 through 7 (nida)  
genes, complete cds; and N-methyltransferase homolog gene, partial  
cds.  
ACCESSION AF016585  
VERSION AF016585.1 GI:2558836  
KEYWORDS Streptomyces caelestis  
SOURCE Streptomyces caelestis  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 41097)  
AUTHORS Kakavas,S.J.; Katz,J. and Stassi,D.  
TITLE Identification and characterization of the niddamycin polyketide  
synthase genes from Streptomyces caelestis  
J. Bacteriol. 179 (23), 7515-7522 (1997)  
JOURNAL 98053867  
MEDLINE 9393718  
PubMed 2 (bases 1 to 41097)  
REFERENCE Kakavas,S. and Stassi,D.  
AUTHORS Direct Submission  
TITLE Submitted (30-JUL-1997) 47p, Abbott Laboratories, 100 Abbott Park  
JOURNAL Road, Abbott Park, IL 60064, USA  
FEATURES Location/Qualifiers  
source 1. 41097  
/organism="Streptomyces caelestis"  
/mol\_type="genomic DNA"  
/strain="NRRL-2821"  
/db\_xref="taxon:36816"  
complement(<1..396)  
/gene="nidi"  
complement(<1..396)  
/gene="nidi"  
/note="ORF6; similar to tyli"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P-450 hydroxylase homolog"  
/protein\_id="AAC46023.1"  
/db\_xref="GI:2558837"  
/translations="MVDSTVGPWELSKDANAKELLDFWFSHNRTHHPVFWDEGRQAWQV  
FPRYDDVLTGNSHPEFFSSDFTEVAPTPPELEMLIPGCTIGALDPPAHGPMKLVSQAF  
TPRMAGOEQRIRVIAELLDRVRGQKTI"  
627..13649  
/gene="nida1"  
627..13649  
/gene="nida1"  
/note="orf1"  
/codon\_start=1  
/transl\_table=11



```

/product="polyketide synthase modules 1 and 2"
/protein_id="AAC46024.1"
/db_xref="GI:2558838"
/translacion="MAGHGDATAQAQAQAEKSEDSDAIAVIGMSRPPGAPGTAEFW
QLSSGADAVVTAADGRRGTTIDAPDFDAFFGMSPREAATDPQORLVLELGEAL
EDAGIVPESLGEASVFGAMNDYATLHHRAGADPTTATGLOHSMANRISYFL
GLRGPVLVDGTGSSSIVAVALESLRGTSIGIALAGVNLVLAEGSSAEMRVAL
SPDGRCHTFEDARANGYRGEGAIVLKLADALADGDVYCVVRGATGNDGGPGL
TVPRAGCEFAARACDQAGTRPADRVFVELHGTGTAGDPVEALGANVYGTGRPAN
EPLLVGSVKNIGHLEGAAGTAGVKAALCHERALPAGLNFETPNPAIFPRLKVV
QTAHAALQPGTGGGELLAGVSAFGMGTNCNVLEETPGGQPAETQADACLSASP
MLLSARSEQARALREHLEDSCADPLDIAYSLATTRFREHAAVPCGPDRL
SSALAAALQAGTPRGVIGSTADGCRALALTGQGAHQPGMGQELVTTDPHFAALDE
VCEELQRCGTONLEWMTDPQDPLDRTEVTQPAALFALQVLTTLTARGTQOHLVL
GHSVGEITAAHLAGVLDLPDAARLITABAHVMGOLPHGCMLSVQAHEHDLQLAHHT
GVEIAAVNGPTCHVLSGPTALEETAQHLRQCNVHVKVSHAFHSLMDPMLGAFR
DTLNTLNYQPTTILISNLQGIADPNHLCTPDYMIHARHTVPADAVQTAHQGTT
TYLGTGPHPTTLTLHHTLDNPTTPTLHREPEPETLTOAIAAVGRTDIDWALVC
GASRPRVELPTYAFQRTHWAPGLTPNHAADRPAAEPQRAVAGVSRGALVRLVG
ETLADVLGDPDEVALDRPTSSGLDSMTAVELAGLGTAGVALDPTLYELPTPR
AVADHAKTLIGESADADQVNGRTGEAEAKADPTIAGVIGCRPGGVPATDILWE
LVASGDAISITPFDGRMDLGLDPTDESTFGKSTVRGGHFDRAQDAEFGLSPR
EATMDPQORLLSTESFALBRAGVPSLRGGRTGVFVTTAPEYGRLLHEGTDYE
GFLTGTGTVASGRIAYALGTRGALPTVDTACSSILVALHLAQSLRGECDLALAG
GTLTVMSGPMFVESRQGLAPDGRCKAFSADADTAWAGVGMILLVERLSDAERLGH
RVLAVRTGANQDSNGLTAPSGPAQOQVIRDALSDAGLSADDIDAVEAHGRTAL
GDTTEAGALLATYGHKQTPVWLGLSKSNIGHTQAAGIAGIKMVALRHDTPPT
LHADPSSKQDMDAGQLLTDARFPWADPDRPRKAGISAFGVSCTNAHVLEBPAP
BEAPARTQAPPVIAPLSANHTPTALRAQARLSMLHHTQADPODIAHALATRTHF
KHRAVITGRTEHLKHTLDAIOGTAHPHRLTLFTGQGAHQMGQELIAYDHPH
FAALDECEELQRCGTONLEWMTDPQDPLDRTEVTQPAALFALQVLTTLTARG
TOAHVLGHSVGEITAAHLAGVLDLPDAARLITABAHVMGOLPHGCMLSVQAHEHDL
DQLAHNGVIEIAVNGPCHVLSGPTALEETAQHLRQCNVHVKVSHAFHSLMD
PMLGAFRDTLNTLQPTTILISNLQGIADPNHLCTPDYMIHARHTVPADAVQI
THTQRTTYLTHGHTLTLHHTLDNPTTPTLHREPEPETLTLATLHTTGGT
ADLRTTYLTHGHTLTLHHTLDNPTTPTLHREPEPETLTLATLHTTGGT
VPSAREWLLGSDPADAILGVEQLVGHGAQVRLDIPAPARGLTLLPSCSELVR
QPSHTEGCGSLRSPRVAEVALIOALGADADPLMATGCAVAAQAGPVSVGA
KWLGLQVAGLEADRWGLVDLPADPAALRGUARLVNSAPNQVAIRASGVFVR
RVPAAPRPVRNDWAPSGTVLITGGTGLGQSVARLALAGAPHLILLAGRNSGAAE
LYDELTALGAETVAACDAADRDLVSLIATPEHRPLTAGVALHAGVLDGGLTTP
ERIDARAKATAARHDELTDADLIDFAPSSIVGVNGGQATYAAANAALDALLA
HGRARGORATSIAMPGWAGSMAAGDGAKSALRDIIVLREPELDELLEOALDET
TLVADWNEFLSSASRFPOLPLFGLPARAAVAVSAGDSDPDPVWLSPOSLSAA
ERQLDLVLRTEASVLEHGSATVADETAFRACGDSLTVELNRLASATGRLS
NTLLDHPPTPAAYSLGELQGLTSPSAASHGADAGARTVASDEFAVWAGVCPG
GYATPDDLWELVAGSDTAISTPFDGRMDLGLDPTDPTGKSVYRGGHFLHDAQF
DAEFFG1SPREATAMPQORLLLETWEALHAGIAPHTLRTTRGVFTGSAQBYGT
OSREAADKYGHLLTGLTASVMSGRVAYTLGQGPALVDTACSSILVALHLAVQSLR
EGECDLAAGSSTVMTPTTVFESRQGLAPDGRCKPADTADGTAWAEGVGLLVE
RLSDAERLGHVLAIVRGTVANQDASNGLTAPSGPAQOQVIRDALSDAGLSADDIDA
VBAHGTGTPGLDPIEABALLATYGHKQTPVWLGLSKSNIGHTQAAGIAGIKMIO
AMHGTLPRLTHVDAPTQVDWEAGSLQLLAEARFPWADPDRPRRAGISAFVSGTNA
HVLEPEAPEAPARTQAPPVIAPLSANHTPTALRAQARLSMLHHTQADPODIA
HALATCTTHKRAVITGRTEHLKHTLDAIOGTAHPHRLTLFTGQGAHQHM
GOELVTHPHFAALDEICEELQRCGTONLEWMTDPQDPLDRTEVTQPALEALOT
ALYRLTTHQOHLVHGSVGEITAAHLAGVLDLPDAARLITABAHVMGOLPHDGM
LSVQAHEHDLQLAHNGVIEIAVNGPCHVLSGPTALEETAQHLRQCNVHVKV
SHAFHSLMDPMLGAFRDTLNTLNYQPTTILISNLQGIADPNHLCTPDYMIHAR
TVRFADAVQTAHQGTTTYLGTGPHPTTLTLHHTLDNPTTPTLHREPEPETLTTA
LATLHTTGGTTPHPSHIAQORVSLPAYFORRAVWMPNSAAHIGSDAEALTLGLA
RODHPTECATSIVSGATLTVGRVSLAAHPWLAHDTWAGVLPCTAFAILLHVAB
ETGSGTVBELTLHVLVLPAQAVQLQVLVEAPDERGRAATVAARPEETGAETHH
AAGVLASGAPPAVADWADGAAMPFGAEAVDVELAEFAAGIDYGAFTGLTG
AVWRGEFADVLPEAVASGEERFVGHFALFDALHPWRVCGPWSQESDADGG
QSTLPLFSQWGLTYGTGAQVRLVPLSDGAFSVAADAPDPAHLDAHLVLRVP
STTEGALPAAPMYHIQWQVDSLVDTRYAVAGDADETAHVLRAFLPDALTATAH
DLAALBAADVAVPTGPAVVALIPGSPPPAPEATSAQDPVRAHLHGLSLVQALIE
SRAALTGLTTLVAVTQAVATQDVPDLAAGALLWGLVRSQAQSEPTGFRFLDVDSPE
DBRACATATALASPEQALRAGVMLPALTPAEAGASAPADAAPTGRMDGVTLLTGC
TALGRVAVHLLARHGVRLMILVSRSGNAPESALRELTVELVHTTSFVSCDLAP
ADVRKAVAGVSDHPHTVGVHTAGVLDGALARTPTERTIDTVLRKADVAHRLHATL
GHPLRAFVLSAAGLGRPGQASAAANALDAPARLREASGLPAVSLAWGLWDRE

```

```

GMAGGLDATALRRRLRRDGIILMPMPADQOALELDDQALTTTYRDPGAPALLVPLLLDSTARLT
AGENGASVPLLSRSLVPSHPRHRTTATSVRDESQSGPGRSARLAAALPAKERMVTLDD
TVSQAVGLGHAASAEIDPPEHIEFDSLATLELRNLRLGLRVLGRLSTPLADPH
TPRKVAWINNELPAANPALGTSBEDPVLGIDOMARALIMWDEADGRADVQRQLVG
LLTALDPAASGADSTRAAATTVARLDEBATDDEIFAFLEQLEQEPDEPLGAPLDERL
DEQ"
13693.19212
/gene="nidA2"
13693.19212
/gene="nidA2"
/note="orf2"
/codon_start=1
/transl_table=11
/product="polyketide synthase module 3"
/protein_id="AAC46025.1"
/db_xref="GI:2558839"
/translacion="MTAENDKIRSYLKATAEHLTKERLAELESARHEPVAVCMAC
RYPGVATPDDLWELVAGSDTAISTPFDGRMDLGLDPTDPTGKSVYRGGHFLH
AAQDAFFG1SPREATAMPQORLLLETWEALHAGIAPHTLRTTRGVFTGMYD
DYGRCPSPLEVGSVVNGSAGSVASGRVAYTLGQGPALTVDTACSSILVALHLA
QSLRRGCDLALAGSSTVMTPTTVFESRQGLAPDGRCKSFADTADTAWAEGVGV
LLVERLSDAERLGHVLAIVRGTVANQDASNGLTAPSGPAQOQVIRDALSDAGLSAD
DIDAVEAHGTTGTPGLDPIEAGALLATYGHKQTPVWLGLSKSNIGHTQAAGIAGI
KMIQAMHHTGTPRLTLHVDHPSSKVDWKAGSLQAQARFPWADPDRPRRAGISAFV
GTNAHAIIEEPTTLEAPVDETTPVPAHTPPIVAPLSAHTPTALRAQARLSMLH
TOADPODIAHALATRTHFKHRAVITGRTEHLKHTLDAIOGTAHPHRLTLFT
GCGAQHGMGQELVTTDPHFAALDEICEELQRCGTONLEWMTDPQDPLDRTEYT
QAPALFALQVLTTLTARGTQOHLVHGSVGEITAAHLAGVLDLPDAARLITABAH
QOLPHGAMLSVQAHEHDLQLAHNGVIEIAVNGPCHVLSGPTALEETAQHLREQ
NYRHTWLVKSHAFHSLMDPMLGAFRDTLNTLNYQPTTILISNLQGIADPNHLCTP
DVIHARHTVPADAVQTAHQGTTTYLGTGPHPTTLTLHHTLDNPTTPTLHREH
PEPETLTLATLHTTHTTTLHTSQSHHLDLPTYPQDRYVWVAPRAAVGLDA
AAGTAVGHPHLLAAVTEPDETTVMTGMSARALPWLADHLVMDRVPAGVATLEIV
LEIGERICPGICELVLEAPLSWGDADAPVWTVVGLSPECTREVSLSRPEDAVW
TRHASRLARAGSEEMSGVFAELSGVMPPGQAEVDISGQVELFAAGVGGGAFRG
VMTDVRPENGHTLVRVTRAQDADGQDDRPQGVASLAAADRGDFVFFVARSVL
LRAWRREDEIFAEVRLPDDHATDAGYGLHALLDAAHPITALDPLDGLDGGHLLPF
RVSATOLPAAPFVPLHRLDWPIIPGFTSTGAGSDPSVLDLTELTAHVLTVREVL
EAVOAHLEPEVPGDGGRLVLRGAGSGADAPAGAAWGLVRSQAQSHPGKRFVLD
VGTDESSRALPAMVAAVGDVDAPQAVRQGRPSVPLVPLTADTEAPEGPGSGAGL
LDGPTDVLITGGTGAALAVTARHVARHARHLLVSRGTDPQAOAEHLASELGAIGA
DVEFFACDTPDRAAVORLLAGIPAEHPLSCVHTTAGVLDGDIISACTPERVAARVLGPK
ADGALHDELTRDRERSVPLVLFSSVSGVIGSTGAGYAAANAFTDALARRADGHP
LDHACWNRMTGTGMAAGLDADATARIIRRSIGAPLDAEMOLFDRALARPEALLVPR
LDSHAGLAAARAGTLPVLRDLAGVPSRAHPAGSEGTATSTGTGSLAEPLARTVA
ERLPLLELVRTEVAVLGHGDPGAVGAESEKDFGSLTAVDLNRLNARTGLRLP
ATLVDFHPTPLALAEILLGLADVEPADGPELAAAFDRLAGLALDADRAARYGAE
RLRSLTALTADPATGTGAQAGPEADGALTVEVTDRLRSASDDELDFLDYDSEFR"
19269.30233
/gene="nidA3"
19269.30233
/gene="nidA3"
/note="orf3"
/codon_start=1
/transl_table=11
/product="polyketide synthase modules 4 and 5"
/protein_id="AAC46026.1"
/db_xref="GI:2558840"
/translacion="MSATNEEKLYRLRAMGDLHSARERLAELESARHEPVAVCMGA
CRYPGVATPDDLWELVAGSDTAISTPFDGRMDLGLDPTDPTGKSVYRGGHFLH
DAAQDAFFG1SPREATAMPQORLLLETWEALHAGIAPHTLRTTRGVFTGMYD
HDYGSHOVGSAAADAGQLGTVASGRVAYTLGQGPALTVDTACSSILVALHLA
VQSLRRGCDLALAGSSTVMTPTTVFESRQGLAPDGRCKFPADTADTAWAEGVGV
LLVERLSDAERLGHVLAIVRGTVANQDASNGLTAPSGPAQOQVIRDALSDAGLTP
DIDAVEAHGTTGTPGLDPIEAGALLATYGHKQTPVWLGLSKSNIGHTQAAGIAGI
IKMVALRHDTPPTLHHTLDNPTTPTLHREPEPETLTOAIAAVGRTDIDWALVC
SGTNAHAIIEEPTTLEAPVDETTPVPAHTPPIVAPLSAHTPTALRAQARLSMLH
STIADSGALHGLAVGRALPHRAVLGLDGAAPDLAALASGESVSDVVTGSAADVRL
VAFVPPGQGAQWAGMGAELLDSSPVFAELACEALPEFVDMSLTDVLRGAPAGPL
DRVVDVQVPTFAVVALAAMWRMLGVEPAAVVGHSQGETAAHVAGVLSLEDAARVA
LRSQILATLAGRGSVAALAAVDSHRSRHEGVRDELAAGVLRAGPRAGRVFYST
PGASVLSLGLQEGVVRVIDVYASHRSRHEGVRDELAAGVLRAGPRAGRVFYST
VEAEPDGLDGTALDAGYWRNLQRVRFESALRAMLADGVDAFVECSPPHVLTVPRQTL

```

gene

CDS

gene

CDS

EDAGAGAVAGSLRRDGGRLRRFLTSAAEAQVAGVPVDMALCPRAGWDLPTVAFOR  
ERYVPAEPGPGAGAAATGPAATAARSGSQDAAASAYCDRLAYHVAWQGLLP  
GADWRPGLLIIVPAGERAEVADSVQAIASPGTVRRVTADPLSTGRGRLALIA  
PVVAGEEPLAGVSLGLCTDGHPEHPGVPAGVTGTLALVQALADLGVAVPLTVTRG

Query Match 5.2%; Score 26; DB 1; Length 41097;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 GACGGCGCTGCAAGGCGTTCGCCGA 167  
|||||

Db 24315 GACGGCGCTGCAAGGCGTTCGCCGA 24340  
|||||

## RESULT 37

AF263912 123580 bp DNA linear BCT 24-MAY-2000  
LOCUS Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,  
DEFINITION complete sequence.  
ACCESSION AF263912  
VERSION AF263912.1 GI:8050835  
KEYWORDS Streptomyces noursei  
SOURCE Streptomyces noursei  
ORGANISM Streptomyces noursei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 123580)  
AUTHORS Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strøm, A.R.,  
Valla, S. and Zotchev, S.B.  
TITLE Biosynthesis of the polyene antifungal antibiotic nystatin in  
Streptomyces noursei ATCC 11455: analysis of the gene cluster and  
deduction of the biosynthetic pathway  
Chem. Biol. 7 (6), 395-403 (2000)

JOURNAL 20334850  
MEDLINE 10873841  
PUBMED  
REFERENCE 2 (bases 1 to 123580)  
AUTHORS Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strøm, A.R.,  
Valla, S. and Zotchev, S.B.  
TITLE Direct Submission  
Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim  
N-7489, Norway

FEATURES  
source  
1. 123580  
/organism="Streptomyces noursei"  
/mol\_type="genomic DNA"  
/strain="ATCC 11455"  
/db\_xref="ATCC:11455"  
/db\_xref="taxon:1971"  
/complement(46..783)  
/gene="nysF"  
/complement(46..783)  
/gene="nysF"  
/function="presumably post-translationally modifies ACP  
domains on PKS"  
/note="putative 4'-phosphopanthetheine transferase"  
/codon\_start=1  
/transl\_table=11  
/product="NysF"  
/protein\_id="AAF71762.1"  
/db\_xref="GI:8050836"  
/translation="MIELILPATVATYAAVDPRPGDRLLSSEREVIARAVSRQRE  
FTVTHLRLALRGLHPDRAILPNRRGAPQPPGIVGSMTHCAGYRAAAVPAELSA  
AVSIDAPNGPLPAGVNLALPSEPHVALAHRPDVHMDRLLFSAKESVFKAWYP  
LTQRELFSAEIVDPQGAFTARILVPGPLLGRRVTVPPGRWHSTPALLITAVHL  
PATPRDRRHRLTWNSLPRPTFG"  
/complement(867..2684)  
/gene="nysG"  
/complement(867..2684)  
/gene="nysG"  
/function="presumably involved in efflux of nystatin"  
/note="putative transporter (ABC family)"  
/codon\_start=1  
/transl\_table=11

gene  
CDS

/product="NysG"  
/protein\_id="AAF71763.1"  
/db\_xref="GI:8050837"  
/translation="MASPDLEERTAPRRLNGLRRHRSVALAVSMVGCVIL  
NAPGILLGRVTDLIADVLGVGPGAPGIDFAAIGRLLVLLALYVASFLEMAQGR  
LVASAVRTHIELRDAREKRLRLPLRHFDRQAGELLSTTIDMLQOTLQOTLAE  
LITISFLTLMVLMVLSPLAVMLLSVPSALIAARTSKRAQHYAAQWSANGTL  
NAHVEVCTGHALIKGFDRAAEERFDACNDAYRAAKAQAQFASGMEVPMFVANL  
GYLVAVIGAWKVIINGTLGLDQVAFILYARQSPQIVETASVAGRLQSGIASRVF  
TLIDPEQADPPRPGTAPRAREGVEFTVDFRSPDTPLEINLSLVEPGSTVAIVG  
PTGAKTTLGNLIRFYEPDSGRILLDGTDTATMTDRSLRFGLVLDQTLFGTITA  
ENIAYAPGACRADIEBARATCADRFITLPOCYDTVLDDESGTSGAGKQLITVAR  
AFLRAVAVLVDLDETSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM  
DAGRIVQSGTHDOLLCAQGIYARLHAARTHTPTAGAAAG"  
/complement(2662..4416)  
/gene="nysH"  
/complement(2662..4416)  
/gene="nysH"  
/function="presumably involved in efflux of nystatin"  
/note="putative transporter (ABC family)"  
/codon\_start=1  
/transl\_table=11  
/product="NysH"  
/protein\_id="AAF71764.1"  
/db\_xref="GI:8050838"  
/translation="MLRLRLAQLRPYANATAALVALQLVQLLTLLPTLGRALIDQ  
GVYRGGRITELGVVMGVVALVQIAAALGAALAAARTATAMGRDLASALFRILDFS  
AREIRGPTSLITRSVNDVQVONLAQTFGVVCAPLMCLGSLVLLAQDVPDPLALL  
LVALLVAVFCGLLARMGTLYARMLTDLRLGLRLREAITGVYRVVRSFVDDHRA  
RFQATNDALFVRSRVRGLIATMLPVLLIMNGFTVALLMTGSHRIDAGMRPGISLA  
LLSYLLIIMVVMVLAFLVFLSVPRARVACRIAEVLDITGSSVAPPAAPQVPGPAGRI  
ELCAAGVYPGAEPVLDVLTVEPGERIAVLGSGTSGKTLINILVRLDAPEGAV  
RVGTDVRELTAATLAAAGVFPQRPVLFSGTVASNLRFGRPDATDELHEALVAQA  
ADFVAMPDGLDAEITCGGNSVSGQRSLRALLRFEITLDFDCFSALDQATDA  
ALRTALVPTAGAVITVAQRIASGRADRIIVLDGRVVAQGHVLLRTSPYREI  
ALSQTEEEAAHGLAGRP"  
/complement(5748)  
/gene="nysDIII"  
/db\_xref="GI:8050839"  
/translation="MSKRALITGITQDGSYLAHLISQGYQVWGLIRGOANPKRSRV  
SRLASELDFIDGLMDQGLSVASVDTVQDVEYNLGAISFVPMWQAEVLVTEVNGM  
VLRMLRAIRMVSGLSRTSVSRGOIRFYDASSEMFGKAAETPQRETLFHPSPYV  
AAKAYGHYITRNYRESFGMYAVSGLMNEHSPRGQEFVTRKISLAVARIKQGLQDL  
ALGNLDAVDWGYAGDYVRAMHLMQDAGDDIVIGTGMHVSVDVAVTAEFVGLNW  
EDTVVDIDPLVRPAEVEVLCAQDRAKQDRLGKVPDVPFTLMRMVMSDLAQVSRNQ  
YGDVLLAANW"  
/complement(34363)  
/gene="nysI"  
/complement(34363)  
/gene="nysI"  
/function="responsible for condensation steps 9 to 14 in  
the nystatin polyketide backbone synthesis"  
/note="polyketide synthase"  
/codon\_start=1  
/transl\_table=11  
/product="NysI"  
/protein\_id="AAF71766.1"  
/db\_xref="GI:8050840"  
/translation="MDNEOKRLDYKLATADLRTRRRVHKLSEAAQEPVAILMTCR  
YPGVRSPEDLWEVWEGHGVTPFTDRGMDLEALAAAPTASGGTHDADPDADF  
GISPRVAMDPQORVILSWEAFERAGIDPTVSGKSGTGVIGMAQYRVGPADG  
ABGQUTMGTVLSGRI SYTFGTGVGPAVTDVTSGLSVAVHLATQALRAGCTLAL  
AGVTINSGPGTPIENGQGLSADGRCSRFGTADGTGWAEGVGLIVLERLDAVRN  
GHEILAVVRGTAVNQDAGSGLTAPNGPSQQVIOQALVNARLAAGIDVVEAHGTCT  
TLGDPVEAQLLATYGNRPADRPRLILGSGVKSLSLNTSHTQAAAGVAGVIKVMWRHGTGL

gene  
CDS

/product="NysI"  
/protein\_id="AAF71766.1"  
/db\_xref="GI:8050840"  
/translation="MDNEOKRLDYKLATADLRTRRRVHKLSEAAQEPVAILMTCR  
YPGVRSPEDLWEVWEGHGVTPFTDRGMDLEALAAAPTASGGTHDADPDADF  
GISPRVAMDPQORVILSWEAFERAGIDPTVSGKSGTGVIGMAQYRVGPADG  
ABGQUTMGTVLSGRI SYTFGTGVGPAVTDVTSGLSVAVHLATQALRAGCTLAL  
AGVTINSGPGTPIENGQGLSADGRCSRFGTADGTGWAEGVGLIVLERLDAVRN  
GHEILAVVRGTAVNQDAGSGLTAPNGPSQQVIOQALVNARLAAGIDVVEAHGTCT  
TLGDPVEAQLLATYGNRPADRPRLILGSGVKSLSLNTSHTQAAAGVAGVIKVMWRHGTGL

PRTHABEPHTHVDSQAVRLTDTTDMPATGAPRAAVSSFGISGNTAHTIIEQAP  
EPOEDATAOQADSTPATAPVVPVVLGGRTDPALRGQAALRAALDTCRRP  
DLDLAHSLATRAGFHRVALLATDPALDGLTALADADDPAAPAWITGTTRAET  
RLAVLFTQGAQRAGRELARPAFPAALDADALDAPHLDRPRLVGTDAALI  
DRITAPALFAVALVRLTSTGFRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA  
ARGMLQALPDGGAMIAVOASEADVAPLHAGHEDQVAIAAVNGSPAVSVLSAEATVTA  
LAEQALADGRTRILRVSHAFHSPLMEPLMDAFRAVEDTLQPLLPVNSLTKGPA  
TVAOLITADYVWVRHVRHAFRAGDIDMLARHDTTAFELBGPGLSAMAQDCLDAADA  
DAVTLPALRACRPREHTLTALAGLHVHGTLDWTGFCAGTCARRTDLPTTAFORRY  
WPKALQSTADLRSVGLGAHHPHLLSAVSLADAGSTLLTGLSRQTHPMLADHVRG  
TTLPLPTGLBELAVAGVEGCDRVEELTLAPELLPQQGVQVQLHINPDVSGRRP  
VNHARPDTGDDTTHATGVTITADASROLPASSEQGTPPLAGDHPALDAQWPP  
AGAEPLDHYDLRLADGFGYFPVQGLRAAMRGDGVYAEVLEPERSGDAEALFLV  
HPALLDAAHAAFTGLGERGGGLFSGWEGVSHAGGATLRLVRLTVPADALATL  
ADGTGAPVLVSIVLRSVATQOLDTAADAVARDALFLDWTVPQATDPGPVALLGA  
DPGLLTHAGPADAPDLAALAAAGVPTTIVLSLAGTCDADPARSAHRCAAE  
ALAAVQTLMDHHERFAARLVFTRGATVGRDVAATAAVGLVRSQASNFCFALVDL  
DPDGAAGAAVLVAALVSGEPQLAVRGDVLRLVRLPTEVAGADGTGOGVGGSG  
VFSGEGAVLVTGGTGLGAVLARHVAEYGVDRLLVSRGSEAGVAGELVAELAGV  
GARVRVACDTRDAAVVELVGGHVASVVAAGVLDGMVGAITGERLSAVLRPKVD  
AVHMLHEATRGDLDADFVFSGLAGVSGPGQAAVAAANFLDALMTRRAEGLPKLS  
LAWGPEQSGGWTGLTDVDAERLARSGVPLSVAQGLALFDAAVAGTDTATCPVRLD  
LPVLARGSEVPLLSLRIRVARRAASVATAGNLAAQLRLDEGDRMDLVLVGR  
QVALVLGHATGVDAGRAFDLGFDSLTVAELNRNLNTVTLRPLATLVDIYFVRH  
LATVYLDLGLTDAEVTPAANAADDPIVIGMACRYPGVSSPEDLRVLEGT  
DAVSGFTNRMGDSVSLYHPDHPGTGYTRSGFLHEAGFDPFGFGMSPREALAT  
SQRLLSYSSWEALIRAGIDPVLSRGTGVAGVMSDYSAMLASPEFPGQSGSS  
PSLASGRVATLGLRGPATVDTACSSILVAMWAOALRSCGGLALAGGVVMTSTP  
AVVDFARQGLSPGRCKAFADADGVWSEGVGVVLRSQSDAVRNGHELIAVRG  
SAVNDQASNGTLPNGSQQRVITQALASGLTADGVVVEAHGTGTGLTGPIBAQA  
LLATYGRDRPERPLLLGSVKSNLGTQAAAGVAVIKVLMARHGVVPTLHVDAQA  
SHVDSMEGAVELLSQAAWPTGVRRAVSGVSGISGNTNHHVIVPEQAPCAKAAAGA  
ARRTPGAVPLLSGRGRLGQAALHGLHQAEPDAELVDVALSTRSFEQRAA  
VVAORDOLASIGLALADRPDPAVEGEAAGRTAVLFTQSGSORAAMGRELHEVO  
SRMDPGLVLLSQAALPQVRRAGVSGISGNTNHHVIVPEQAPCAKAAAGA  
PEFAAFRAVCAVPLDLRLEVRVFAEDGSDAALLDETGTQTPALFAVEVALFRL  
VESWGRDPFVAGHSIGETAAHAAVAGTLDLACRLVARATLMQALPTGGHAIQAIQ  
TMDTAHLDLDTVAIAVNGVQSVISGDEEAETIATFAERGRKTRKRLRVSHAFHS  
PRMDGMLDAFRITVAEGLYTRAPILPLVSLTGRADDAEVCVTAAYVVRVREAVRPA  
CVRTRDAGATFELBSGDLTMAEDTGLDHDHAEIVPMLRAGRAELAAATALAR  
LQVTRGVDDWAAVLAGTGARETDLPTAFQHAIVYVLPQTPPAAALAAADPADQOLWAA  
VERGDARELILIGEQDLTDFDSILLPALTSSWRGNQKHLDLTRVREVTRLSKP  
TAPLDVGTWLLVASDADATDALLDGLADALSHGARVRLLLDDSDCADRAVLAERL  
ARTADVDAATVLSVLPDERRDADCPPLTRGLALTVALVQALDQAGRLMTATRG  
AVSTNPADPVTPOVAAAGLGRVLAHPRLPGLVDPQVDFERAGORLAGILAVK  
DAPDCEQVALRATVSGRRLVTRVEALPTAAEFATGTGLTGTGGLGAELVARWL  
ABAGAQLVLTSEKGPDAACAELEAELEGGYSPSVVACDVADRDALAAVLTALPEE  
LEPLTVHVGHTVGLPDLTSTAEPAGLTAALAGAAHLDALDARELDLFFVLPGSI  
AGVWGSQSAAGAAAYLUDALHRRHRAGLAATSVAMGPWABAGAADDAVSETLRR  
QGLGLDPAAPANTELRVAVRQDVTTVADVDMQRYAPLTSARPALTAGLPEVRAL  
AADERTQDAGTGAESVTVRRALAEPEQLRLTDLVLTESATVLSHSSADAVPEGRAL  
RDVGFDLTAVELRKLGAATGLSLPSWVVDYPTLELAQVLRABILGAELVAGVP  
ATGQADDEPIALIGMACRFGVSPQOLWLVASTDAISEFPVNRGQWGLHLPDP  
PDRGTYTQGGFLHEDFPTFGISPREALVMDPQORLLLETWESPERAGIRP  
ETLRSLTFTFGSSYQYEGSGLAGDGTGHTVSPSVLSGRSLSVFLEGPATVD  
TACSSLLVHLACSLRNGENLAVAGGATMTTNPFIASFQRALAKDGRCKAFS  
DDAGMTLAEVGVVLREAEQNRHPLVAVLRGAINQDASNGLTAPNGPSQOR  
VIRQALANARLAPDIDALEAGTGTPLGDTAQALFATYGRDRDPESALLGLSVKS  
NIGHTQSAAGIASIVIKWVALRHELPPLTHADAPSSHVDMWNGTVLLTQRAWDET  
GPRRAASFGISGNTNHHVLEQAPVADTPAEERPAVAPVPTAGVFWVPTVARSAA  
ALRGQRARLLAETGTLPAGPLDILGLSVARSARFHRVAVVPPVPLALR  
AVATDGPVSVAGVADVGRVTVFPFGQSGQWNGMSQLLDESAYFAERIAACRAAL  
AEFTDMSVLDVLRGVGASLERVDVVPASFAVMVSLAALWRSQRALPDAVHSHQ  
ETAAVVSGLALRDGARVVALRSQAIGRALAGRGMSVALSVSHVLEPVEFEGRV  
SVAVNGSPSVRGAPEALDALHATDADIRARLAVDYASHDVLEHLEELV  
LAEAPRTSPVFTTGDMLDTRMDADIRARLAVDYASHDVLEHLEELV  
EVSHPVLSMAQVQAEAGVPAAPAGTLRDQGGTDRFLLSAAEVGRGVDDWAGL  
FEGTGASRLDLPYATQHEHMAVPAPEAAADDDAAFTWAGDVGDVSAULTAALG  
TDEDSVAVALPALTWRKARRDRSTVDWARYRVAWPKGLGTTPHPSLTLWLTADG  
IDTDSVALETVGAERVLVDEECVDRVLEERLAGAEDVTGIVSVLAAABRTDVA  
PGTSLVLTGALTVALQALGDAIDAPVWALTGAVSTGDAELTAPVQVGTIGWT

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 142 GACGGCCGCTGCAAGGCGCTTCGCCGA 167  
|||||  
Db 51497 GACGGCCGCTGCAAGGCGCTTCGCCGA 51522  
|||||  
RESULT 38  
LOCUS AX2111739 125401 bp DNA linear PAT 06-SEP-2001  
DEFINITION Sequence 35 from Patent WO0159126.  
ACCESSION AX2111739  
VERSION AX2111739.1 GI:15523950  
KEYWORDS  
SOURCE Streptomyces noursei  
ORGANISM Streptomyces noursei  
REFERENCE 1  
AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,B., Brautaset,T.,  
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and  
Gulliksen,O.M.  
TITLE Gene cluster encoding a nystatin polyketide synthase and its  
manipulation and utility  
JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;  
Norres Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN  
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;  
ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich  
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;  
Brautaset, Trygve (NO) ; Sirom, Arne Reidar (NO) ; Valla, Svein  
(NO)  
FEATURES  
source Location/Qualifiers  
1..125401  
/organism="Streptomyces noursei"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1971"  
/note="AFCC 11455"  
ORIGIN  
Query Match 5.2%; Score 26; DB 6; Length 125401;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 142 GACGGCCGCTGCAAGGCGCTTCGCCGA 167  
|||||  
Db 51905 GACGGCCGCTGCAAGGCGCTTCGCCGA 51930  
|||||  
RESULT 39  
LOCUS AF521085 132544 bp DNA linear BCT 23-MAY-2003  
DEFINITION Streptomyces nanchangensis NS3226 polyether ionophore nanchangmycin  
biosynthetic gene cluster, complete sequence.  
ACCESSION AF521085  
VERSION AF521085.1 GI:31044135  
KEYWORDS  
SOURCE Streptomyces nanchangensis  
ORGANISM Streptomyces nanchangensis  
REFERENCE 1  
AUTHORS Sun,Y., Zhou,X., Dong,H., Tu,G., Wang,M., Wang,B. and Deng,Z.  
TITLE A Complete Gene Cluster from Streptomyces nanchangensis NS3226  
Encoding Biosynthesis of the Polyether Ionophore Nanchangmycin  
JOURNAL Chem. Biol. 10, 431-441 (2003)  
REFERENCE 2 (bases 1 to 132544)  
AUTHORS Sun,Y., Zhou,X., Dong,H., Tu,G., Wang,M., Wang,B. and Deng,Z.  
TITLE Direct Submission  
JOURNAL Submitted (13-JUN-2002) Bio-X Life Science Research Center,  
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,  
China  
FEATURES  
source Location/Qualifiers  
1..132544

Query Match 5.2%; Score 26; DB 1; Length 123580;  
Best Local Similarity 100.0%; Pred. No. 0.15;

```

/organism="Streptomyces nanchangensis"
/mol_type="genomic DNA"
/strain="NS3226"
/db_xref="taxon:204925"
complement(682..2052)
/gene="nanT1"
/gene="nanT1"
/feature="putative integral membrane transport protein"
/codon_start=1
/transl_table=11
/product="NanT1"
/protein_id="AAP42848.1"
/db_xref="GI:31044136"
/translation="MHOTTDDPRASRRSGVRRFAVSLTGTALEWYDFAVYAAAA
LVFGDLFPFEDPLTGLLAFSTYAGVGVSRPLGIVFGRIGDVIGRKKVLIALVLV
GAATFLGLVPTSTYIWAAPIALVIRFAGGVGVGGWGANVLLSEFGDSRRRGFY
ASAAQVGGPNGLANGVLAAGALLTEAQEAAGWRVAFLLSGALVAFGLIRAKLE
ETPVKAEAPQSRPEAPIREVFTQPRALLAAILCRVPDVLVAFVFTVLTATGE
LMSRGALAAVLIGSSIQVFLMPLAGAVSDRIINRVLYGCAVAAGVWVPLFFPMIG
GGSVPLAGVVGVLVTHSFLYGPQAAFIAQFSPRLRYTGSSLAYTLAGIIGAIAP
LFTLLLSAGSWLPLALYIAVAASVLAGVLLGRDPEAAVDEDAQLATPAQAATADA
ARPTAV"
2064..2879
/gene="nanT2"
2064..2879
/gene="nanT2"
/feature="putative ABC transporter"
/codon_start=1
/transl_table=11
/product="NanT2"
/protein_id="AAP42849.1"
/db_xref="GI:31044137"
/translation="WLGHRPSGILLVRSYAQALRKLTAEKSTTADRHNSGSAPM
ARLTFEIPDGSTVEDIVQVILNAGVAGSODDVAHIAELAEGLVPTSPVYALYVVA
PYLAQADRVAQHRRTSGEAEALVAVAGDGELLTACDHTDRDLELVHGVAVAKNAG
PDVLAIRLADVEPRDLDTLRAWTRDGTETEIQHGTIAELLTPAYVVDVLRSD
ALTPGTVLISGTIPMTPGVDQFADTWIRVELGDPATGDTIRLAYDVHPMPEPIG"
3117..3818
/gene="nanT3"
3117..3818
/gene="nanT3"
/feature="putative two-component response regulator"
/codon_start=1
/transl_table=11
/product="NanT3"
/protein_id="AAP42850.1"
/db_xref="GI:31044138"
/translation="MSTRCTGSLTNGVSVBELGIPLSVVLVDDHPVVRAGISAWCAAA
DAPISVVAEGANVSVALHGPGRGADVVVMDILLQNGRPAYDELQELVAKERVVYTM
RDSQDAALTCMDLGSATYITKAEQORHLVKAIRAAEDIPYTPSLAGAFGSDTRQSR
PVLVSREVELVWFQESKAVVAQSLGISERTVNTYLDVRVIRKYANAGRPAATTAKL
VARAVQGLIALDEL"
complement(3724..4923)
/gene="nanT4"
complement(3724..4923)
/gene="nanT4"
/feature="hypothetical chemoreceptor"
/codon_start=1
/transl_table=11
/product="NanT4"
/protein_id="AAP42851.1"
/db_xref="GI:31044139"
/translation="MGEDFROLITPHQLRGIRAGGLVLTVLTHIGSALPGLLGIKEY
RHPPIPLAAYGLLTLILGSSVVLGLDGRPWGRWPCALAGTFVASVATTQQVPADHY
FVSSVSLVSLAGNVVLLHGRPLVSGAFGLAHVLAALLLVDPVPSISGASMA
LSASAGCFQMTIAGAKLLDSSAAIQALRAQSRVTRIAVAFQIQDQRRYAEI
NARNVTLTGLASGLDDEDDHACLEAARLRLEAAGTSDCTFDPIQVHEMRACIDV
ABERNVTQLAVRGDPLEPLPLRLRALDPLVLAALAAQOARTVTVRGDQVGVV
VDALDRLEPEPKANGIRVTRVHAEDRLVVEACRITARPPELSDRPEPEPWPANPSWP
AGRSRT"
complement(4916..6106)
/gene="nanT5"

```

---

```

complement(4916..6106)
/gene="nanT5"
/feature="putative two-component sensor histidine kinase"
/codon_start=1
/transl_table=11
/product="NanT5"
/protein_id="AAP42852.1"
/db_xref="GI:31044140"
/translation="MSISSISPGIDPPASARLPGSAGPWAREYVAFNRRLIYGRAGV
MSCAVLGVAAMPMEGEMAPAAVAAVLAWSWIHLRLATTSAMQRPILLALDLTVMTA
LCLSQRTIPEAQTHAGTHTLAVSVTAVALQMQPPLTGILLATLCCADVVTGTA
HPNLESGVLRSVIWWVNTALARAVALRVLFVSESRADRAARARQOGEAEARAC
AAERHIALHDHTACATLLIASAPWASRAETLRAQAARDLLRURASOPVAGEVLLAA
ELDDEIAHLPLRVINREFDELGTPTFAALRGGLGEALRNVARVAGVAVRTAER
VDHVIVTTIADSGVGFDEHPHIGRIKRSIHGRMNVVGVRATVESRPGHGTTRLE
WPRG"
complement(6768..7493)
/gene="nanR1"
complement(6768..7493)
/gene="nanR1"
/feature="putative regulatory protein"
/codon_start=1
/transl_table=11
/product="NanR1"
/protein_id="AAP42853.1"
/db_xref="GI:31044141"
/translation="MGPLRVVDEKSGSPRAQKLAALVALLIRSDOVVMAEQUTREI
WGELPPRRDTASIVYCI SRLKFLSRPQQNPITRPGVYMLRRASCECDPFDVQQLV
ADGKCYVGEPAQALCFEQALALWGPAGNVQTCPIVAGFLKRLDEMIECAEMS
AEAHLEGRRELVGRLRSIVAEFPRTETFRQLMLALYRSDQSEALGYESARRVL
DRELGLEPGRKRLRELHRTVLAAT"
complement(7573..8334)
/gene="nanR2"
complement(7573..8334)
/gene="nanR2"
/feature="putative regulatory protein"
/codon_start=1
/transl_table=11
/product="NanR2"
/protein_id="AAP42854.1"
/db_xref="GI:31044142"
/translation="MRYEILGLRVDRSSVTSI SARKIEVLVLAALLVRADVVSSD
LVKEIMTRAPQRAVALHYVSQLRKFLRPNETSPITVARSYVLRIGDDELDFH
VFQGLVRGAAQAGRAQTHAEACAEYEEFWRGFEVLEDLRDAIVNGFVCAEQARL
ECVESFVAGLELGRHEFVSYLYARIEELPLNEAFYQLMLALYRSCRADALHYQ
SARAVLSEELGLEPSAALKRLHHAVLVDAPLEAA"
8919..17627
/gene="nanA1"
8919..17627
/gene="nanA1"
/feature="PKS"
/codon_start=1
/transl_table=11
/product="NanA1"
/protein_id="AAP42855.1"
/db_xref="GI:31044143"
/translation="MRFYGVWAGSSPSHAQATSPVAIVGLACRLFGAPDPEAFWR
LLRAGANVVPVDSRLPTEGPPYFAGLLEHVDYTDGFFGISPRATAMDPOQLR
MLELWAMEDAGLGPKNLAEERTAVFTGATWDDYATLLHRRRNDIALTRHTWAGHL
RGLIANRVSHLLGLRGPSLTVDAQSSGLVAVHACESLREGEADLALAGCVNLIAE
ESMMAEAQFQGLSFGDCYTFDARANGFVGGGWWLLKPLAAAVADGDPVYVIE
GSVNDGATDGLTRPSDQATDVVRQAQRAVSPAELOVLELHGTGTVPDGPDEIAA
ALGDALGDGCHRDRAVDAPLRGVKTVNHLGAAGIAGLKTALSIIHRRLLP
SLNFTAPGIPLAELGLRVCTAFGMPDERRLTAGVSPFGMGTINCHVVLARPPAP
AVLPDRTPAGDVSARTAPVPMVVSASPKALTAQAALEYHLRAHPGLHPVDIGH
ALATRTAFPRAAVVLGRDEDELVRSLDALSETQTSQVINGRAGGRVAFILFSGQS
QRPMGRLAEVAFVADALREVCALHLDPLMDTDTPLDLMFAPAPDGPPEPLNKTAI
TPQALFAIYVALYVLTWSVTPDHLMGHVSGETTAARFAGVLSLPDACTTVAARGRL
MOSITAPCAMAAQOATAEAGQALAEYGRVGLAAVNAVSWISGPREAAVEATEAW
PARCKATLVKVSHPHSLDGLDGLTRTVAAGLTFAAPAI PVVSNLTGGAATEAOL
RSPDIYADHARQAVRFDSAGVHLCDAGVDTFLELGPDSLGTGASLGTGASAAAGDAPRP
VAVAVQRREPEAQSFVSMAQAQVGVGVDAFAAFAGHETRRAPLPTIYAFQDRHWP
DGLDERGARRESTSPVVFESTDREPVVVDASPADRAASPGELLALVTRTHAALVLGHNSP

```

DGIDPALTFKQGFDSLAATELSERLSAATDTTELPATLTFDHPPTNVAAWLRAAHEG  
OPTAAPTAACTGSAEDPVAVAVACRYPGVGESEALMLRVLDVGVDAVGFPGDGRW  
DLAELFGAPDGSAGSAGTGGFLYGAGDFDAEPFGISPREALAMDPOORILLLELSWE  
LLERAGIIPASLASATGTVYGCATANDYQGRLEHATAEILDGHLTGSTPVSASERVAY  
ALUGLEGPAITVDVTTACSSLVNHLAAQROGECODLALAGVTVNATGPMFTFSRQR  
GLAPDNGKSPQAAADGTGMSSEGAGLVLLERLSDARRNGHQVLAVIRGSAVNQDGSN  
GLSPDNGKSPQAAADGTGMSSEGAGLVLLERLSDARRNGHQVLAVIRGSAVNQDGSN  
DDEPLWLSIKNSHTQOAAAGVAVIKMVALRHGFLPASHIDIDAPSHIDWSDGTVP  
RLLESPDMPCTDMPGSDRPRRAVSSFGISGTNAHLILEOAPDHPPEPTTSGGVTV  
WYLSARTADALRAQAGLAEWWTAGAPRSPASPTSPADYVGNHSLATRSADRHRV  
VSGTRDELLEGLRAVADGAPAAVSAAGAPGVNVPFGQSGWRGMSVELLSDSPVF  
AARMAACEALGFLVSLTAVLRGAPGAPSPSRVDVLOPCLWAVMVSLLAWESYGV  
TFTVADGSGSEIATAACVAGSLGSDGAVVLRASQALRALAGHGTWASLALSGAEAE  
RFLADLGAARVTAACVAGSVTVSGPTDQVAAVAAACEAGHRAARTIDVYASHG  
POVRLADTITDLDLSPGASDAVYSAVTEGARQPTTELDADYFNTLROPVAFASA  
IDALLAGYRVFIEVSHIPVLLIPALRECFEAEAEAAATVPTLRDQGGDPQVAAALGD  
GFVAGLAWDRSGVFGDREAGDEHRTVELPTYPQRRRYWLPADPHGRREGRTAG  
VGTTPAGHALSSAVELADGGLVLSGRLPDGAAGVAGTAVGQVLPVAGVLDVWALLA  
ADEAGASLEELLRLAPLELSPGLSPGLSVLAQVAGAPDPSGRRELRSRPPAD  
AGAGEWTCVAVGLAPGAPADTGTATVWPAGABADPAGLYERARRGVYGV  
PALRGVVALWDGADLVADVALPSEAGGGGEGGADGCTAGFGLHPVLLDALOPALI  
AEPDGTGEGGAPARLWLPFAMSGVRLMATGARARVRLSPDGGGGDVADERELRI  
EVSPTDAPVLSVSVLPRTRQVREASGAAGGLFALDWTVPVAPPSGAEDDAG  
CVNLGGEATPEFVGCDTDTYDLPALLAALDAGAPESVVMVMPAADPGAAEDAA  
LSAVRGVAAALRAVWAEPLRTVSLVAVTRGAVAGAGGEPFVDLAAAAGWCGARGVQ  
AEHPDRIVLVDVDDVDMDGADTDTDIGAAGLAAALGPQVALRGDTLLAPLRSAA  
TPGGVAFDPTGTVLVDSDGSLAGSVAEHLVRAEGVRLHLLVPRGAGVAYTVDRQD  
AQVHWTVDPDRTAALERVAVQDPAHPLTGVVHVAGLSADIESGARGHVAAYGVV  
LHLCQATLALSVFRTVLSDAATADGPAAPERAAGAFCAATVDVRRRAGLHGDVVA  
FGPMAAADGGAGSGGRTVGLGADRGALLRACRADRPRLVAAADIRITALTAHPA  
HELPAALRTGASASASAGRAPVRRVAAAAPGRTTWDWASRLVGLGPAERRVLELV

Query Match 5.2%; Score 26; DB 1; Length 132544;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 AGGCGCTGCTGCGCCACTACGCCAC 468  
|||||  
Db 75927 AGGCGCTGCTGCGCCACTACGCCAC 75902  
|||||

RESULT 40  
SEERYA  
LOCUS SEERYA 9769 bp DNA linear BCT 14-OCT-1991  
DEFINITION S. erythraea eryA gene for 6-deoxyerythronolide B synthase.  
ACCESSION X56107  
VERSION X56107.1 GI:46975  
KEYWORDS eryA gene; erythromycin.  
SOURCE Saccharopolyspora erythraea  
ORGANISM Bacteria; Actinobacteria; Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.

REFERENCE  
1 Cortes, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J., and Leadlay, P.F.  
An unusually large multifunctional polypeptide in the  
erythromycin-producing polyketide synthase of Saccharopolyspora  
erythraea  
Nature 348 (6297), 176-178 (1990)

JOURNAL  
MEDLINE 91043075  
PUBMED 2234082  
REFERENCE 2 (bases 1 to 9769)  
AUTHORS Leadlay, P.F.  
TITLE Direct Submission  
JOURNAL Submitted (24-SEP-1990) Leadlay P.F., Dept. of Biochemistry,  
University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK  
FEATURES  
Location/Qualifiers  
1. .9769  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/strain="NRRL 2338 (sub-species)." /db\_xref="taxon:1836"  
57. .9593  
/gene="eryA"

CDS

57. .9593  
/gene="eryA"  
/codon\_start=1  
/transl\_table=11  
/product="6-deoxyerythronolide B synthase"  
/protein\_id="CAAJ9583.1"  
/db\_xref="GI:46976"  
/db\_xref="GOA:Q03133"  
/db\_xref="SWISS-PROT:Q03133"  
/translation="MSGDNGMTEELRYLAKRTVTELDSTVTLRLEVEHRAAGEPTAIV  
GMACRFDVDSPSPFMEFVSGGDAIAEAPADRWEDDPDARLGGLMLAAGDDPAGF  
FGISPREALAMDPOORIMLEISWEALRAGHDVPSLRGATSGVFTGVGTVDYGPDPDE  
APDELGVGTGTASVAGSRVAYCLGLEGPAMTVDTACSSGLTALHMLAMESLRDEC  
GIALAGVTVGTSTASVAGSRVAYCLGLEGPAMTVDTACSSGLTALHMLAMESLRDEC  
ARRRGRPLVLAGSNQDCANGLTAPSGAQORVIRRLAENAGVAGVADYVVEAH  
GTCTELGDPIDVHALLSTYGAEEDDPDLATVSGVSKNIGHTQAAGVAGVAKVIALR  
HGMERTUHFDPSPQIEMDLCANVSVSQAASWPAERPRRAGVSFSGISGTHAVIV  
EEAPEADPEPADSGPVLVLSGRDEQMAQAGRLADHLAASRGTRCATVSRWPP  
AAAEQWVGVGRDEALAGRAVADRTRTADTAATGQGNPRRVRAMVFFGQAQWGM  
ARDLLRSEGVFADSTRICERALAPHDMSLTDLISGARPLDRVDRVQVPAFVWVSLA  
ALWESHGVEPAAVVHSGQETAAAHVAGALTLEDAAKLVAVRSRVLRLGGCGMASF  
GLCTEQAERIGRFAGALSIASVNGPVSVVVAGSGPLDELLAECBAEHKARIIPV  
DYASHSPRESRELLTELAGISPVSDAVLITITGQIDITAMDTAYTANLREQ  
VRFQDARTQLEABAGDFAEVSPHPVLTVEATLSALPADAGACVCTLRDGGGL  
ADFTALGAEVAGQVDEWSPAFADARVELVFPFQELPVYFQFQRYWLPITGG  
RARDEDDMQYQVVMREAESLAGRLAVLTVGPVPSLSDARLSGEOGATVLT  
CDVRSRTIGTALRAADTDLSTVGVAAVPHGEAVDPISDLALVQALGAAGVAPLW  
VLTRNAVQAGELVDPAQMVGGLGVVGLQPCRGWGLVDLDADAASIRSLAALV  
ADPRGEQVIRADGIKVARLVPAAPARATHPLEPLAGTVLTGSGIGHALMARSG  
AEHLVLLGRGADAPGASSELREELTALGTGVTIAACDVADRARLSEVLAEEAASGR  
VSAMVHAAGVSTSTPDLDTLEAETFEIADVVRGTVNDELCPDLDAFVLFSSNAGW  
CGSPLASVAAAANFLDGFARARSEGAPVTSIANGLMWAGMAGDEGEYLRASQGLRA  
MDPRVABELHITLDGQTSVVDMDRRRVELFTAAHRPLFLEIAGARAEQSE  
EGPALAQLAALICDGRERELHLIRAEVAALVGHGDDAAIDRORALVRVPPKDA  
VDLNRLLAATVGVREAAVTFDHPITIRLADHYLELVGAEEAOAPALVRVPPKDA  
DDPTAIVGMACRFPFGVNGPGLMEFIVGGDVAATEMTIDRGWDLDLALDPDPRHGT  
DYSRFGAFDGAADFAAFVFGISPREALAMDPOORILETVELTVELFENDELPHSVRGS  
DTYVGLGAYQYQDVAFPEDSEGLITGNSAVSVSVAVYVLEGGPVTVDACS  
SSVLHSAACSGSLRDCGLVAGGVSVWAGVEVFTFSROGGLAVDRCKAFSAEAD  
FGFLPEGVAVVQLQSLDGPAGCGGVGVVAGSAINQDGTNGLAAPSGVAQORVIR  
KAWAPAGITGADVAVVEAHGTGTRLDGPVEASALLATYKSKSGSPVLLSGVSKNIG  
HAQAAAGVAGVVKVGLNRLGLVPMLCRGERSPLIWSGGVSELAEEAVPMPAADG  
VRAGVAGSVGTNAHVIIIAEPPEPEPLPEFPGVGLAANSVPLLSARTETLAA  
QARLESADVDDSVPLTALASALATGRALHPRRAALLAGDHEOLQVLSRAGVSAEVL  
ATTGTASAGVVFVPGQAGWEGMARGLLSVVFAESIAECDVLSRAGVSAEVL  
EQRPDAPSLERVDVQVPLFVMSVSLARLWCGVSPSAVIGHSGEIAAAVAVGVLS  
LEDGVVRLAKARAKALAGKGMVSLAOPGERARALTIAPWEDRISVAANVSPSVV  
SGDPEALAEVLRCEDEGVRKTLVDYASHRHHVEEIRETILADLDGLSARRAALPL  
ADAVAIGSLHRTABEHLIALARAHVGVAVDMENVPAPFPVALPNYPFPQRYWL  
APEVSDOLADSRVDRWPLATTVDVLSGGFLVHGSAPESLTSVEKAGVVPVVASAD  
REAAALAEVGEVAGVLSVHTGAANALALHQSIGEAGVRLAPLVSRAVALGESE  
VDPDQAMVGLGRVNGLETPERWGLVDLPAPAPGGEAPVACLGDGHEDDOVIR  
DHARYRLVRAPLITRESSWEPAGTALVTGTTGALGHVHARHARCGVEDVILVSR  
GVDPAAALEAEIIVAGLQKTTITACDVADREQLSKLEELRGQRPVTVVHTVGP  
ESRPLHTEIGELSVCAKVTGARDILDELCPDAETVLFSSGAGVWGSANGLAYSANA  
YLDALHRAVRKAVERTSVAMGAWGEGMTGLEGLTFRGLRPMAPNAPRALHQAAL  
DNGTCTVSIADVDMFAVGTAAAPRPLDDELVTFAVGAVPVAPVAAAREMTSEELL  
FTTHSHAAIIGQSDSGTPAREASALDRGQVAGSVRSVLDLGLALPATLIVFEH  
PTVRLADHIQGLSDSGTPAREASALDRGQVAGSVRSVLDLGLALPATLIVFEH  
GSDGFLSDVMDVMDQGEVTVICCATGTAISGHEFTRLAGLRGIAVPRVYGPYE  
EGEPLSPMAAFAVQADAVIRTCQDKPFPVAGHAGALMAYALATELIDRHPRGV  
VLIDVYPRQHODAMNMLELTATLFDRETVMDDTLTALGAYDLRTGOMRPTGL  
PTLLVSAGEPMGPFDDSNKKTWPEHDTVAVPGDHTVQVEHADAIARHIDAWLGGG  
NS"

57. .9590  
/gene="eryA"  
/product="6-deoxyerythronolide B synthase"

mat\_peptide

ORIGIN

Query Match 5.0%; Score 25; DB 1; Length 9769;  
Best Local Similarity 100.0%; Pred. No. 0.57;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```
QY 329 CCGACGAGCGGTGATCCGACGGC 353
LOCUS |||||
Db 5389 CCGACGAGCGGTGATCCGACGGC 5413

RESULT 41
AF220951
DEFINITION Streptomyces antibioticus 8,8a-deoxyoleandolide synthase 1 (oleA1)
ACCESSION AF220951
KEYWORDS and 8,8a-deoxyoleandolide synthase 2 (oleAII) genes, complete cds.
SOURCE AF220951.1 GI:9049534
ORGANISM Streptomyces antibioticus
Bacteria; Actinobacteria; Actinomycetales; Streptomyces.
REFERENCE 1 (bases 1 to 24568)
AUTHORS Shah, S., Xue, Q., Tang, L., Carney, J.R., Betlach, M. and McDaniel, R.
TITLE Cloning, characterization and heterologous expression of a
polyketide synthase and P-450 oxidase involved in the biosynthesis
of the antibiotic oleandomycin
J. Antibiot. 53 (5), 502-508 (2000)
MEDLINE 20363406
PUBMED 10908114
REFERENCE 2 (bases 1 to 24568)
AUTHORS Shah, S., Betlach, M. and McDaniel, R.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-2000) Korean Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
source
location/Qualifiers
1..24568
/organism="Streptomyces antibioticus"
/mol_type="genomic DNA"
/db_xref="taxon:1890"
623..113075
/gene="oleA1"
623..113075
/gene="oleA1"
/feature="polyketide synthase"
/codon_start=1
/transl_table=11
/product="8,8a-deoxyoleandolide synthase 1"
/protein_id="AAF82408.1"
/db_xref="GI:9049535"
/translation="MHVPEENGHSIAIVGIACRLPGSATQEFWRLLADSADALDEP
PAGRPPTGSLSPAPRGGLDSIDTFADFNISPREAGVLDPOORLALBELWEALE
DAGIVPRLRIGRTSVFMGAMWDDYAHLAHARGAALTRHSLTGTGRGMIANRLSYAL
LQGLSVLTDVDTGSSSLAAVHMWDCSLARGESDLALGVGNLVDPPAGTTGVERFAL
SPDGRCTYDFSRANGYARGEGVVVILKPTHRALADGTVYCEILGSALNDGATGEGL
TVPSARAQADVLRQAWERARVAPTDVQVELHGTGTAGDPVEAEGLTALGTARPAE
APLIVGSKVNIHLEGAAGTGLKTVLSIKRHLPASLNFPTSPNPRIDLDALRLV
HTAGVPSPDRPLVAGVSSFGMGTCNVLSLNRNAGGCGAGRPYTGTEDELIGAT
BAEKRPDPATGNDPQADTHRYPLIILSRASDAALRAQAEIRHHLHSHSPGRLRDT
AYSLATRQVFERHVVTHGREDLNLGRLENGLPAPQVLLGRTPTPEPGGLAFLEF
SQGQSPQMGRKLQVPPFRDALDEVCSELTHGLRLGPEAGPPLRDLRVHVRSH
AHSLLSTHYTOALPALETALPRLLVQWGLKPDHLAGHVGHTAAAHAGILDLSD
AELVATRGALMRSLPGGVMLSVQAPESVAPLILGREAVGLAANVPDPAVVSSE
RGHVAALEQLRDRGRKRYLRVSHAFSPLEMEVLEFAEAVAGLTFRATPTLVSN
LTGAPVDRTWATPAVVRHREAVRFGDGRIRALGKLGTSFLEVPDGVLTNARAC
VTAAPEPHRGEQADAHATALLPALRRGRDEARSLTEAVARHLHGVPMWNTSVL
GGDVSRLPTAYAFQRESHWLPSPGEAHPADDTESGTGEAPRPHDVLHVRSH
AAVILGHSAERIDPDRAFDLGLDLSLALRDLRDLTALGLRPLSSVLFPHSPGAL
ARELQGDTRPEPKNTGTRATRGFFLYDAGEFDAAFFGISPREALVMDPQRLLEL
AGEDAVSFPDTRGNVYDSATRGFFLYDAGEFDAAFFGISPREALVMDPQRLLEL
TSWALERAGVPSGLRSDTAIVYATAQDYGPRLHESDDSGVYLTGNTASVAG
RIAYSLGDEGKAVTDTACSSSLVALHLAVQALRRGECSLAGATVWSPGCMFVEF
SROGLSLDEGRCFAATADGTGWAEGVGLVERLSDRRLHVRHLAVVPSGAVNQD
GASNLGTAPNGPSQOVRVRAALADAGLVPADVVEAHGTCTGLGDPLEAQLATYQ
QGRAGRPVNLGSKVSNIGHTQAAAGVAGVMKVLALGRGVVPTLHVDERSHVDWS
AGEVELAVEAVPWRGGRVRRAGVSSFGISGNTAHVIREAPAPPEPEPEPGSGVVG
VVPVWSGRDAGALRQEARLAARHVGVSVDVVGMSLVATRSVFHRVAVMVMSGELAM
```







ALRAGECSLALAGGATVINTPVMVFVFSRQGLSPDRCRSPAADADGTGWSEGVGL  
VLERLSDDRRHGHOVLAVVRGSAVNDGASNGLTAPNGPSQORVIRRALTAARLTAD  
VDVAEHGTGTGLDPIEAOALLATYGOGRPADRLPLWGLSKNSIGHAQAAGVGVSI  
KTVMALRGVLPOTLHVDPAPTEVDMSAGEVPLLTQRREWPAGRPRAAGVSAFISG  
TNAHLLHEAPEAEEDADRAPHLRGMPVWSGRQRALRAARLDRWAAGHPDVA  
PKVGLALATRSVEHTAVITGADHELLSGLSALSGRPPAGVRLTADPAAGGLAFA  
FTGOAQRPMGRGIYETPAYAEAFDECAALDPHLERPLASVVPADPAAGGLAFA  
TAYAOPLFAVETALFRMQSGVHPDLILIGHSVGELSAHAAGVLSIQDAARLVAAAR  
GRMOSLPEDGRMLAIQASEDEVLPVAFELSEKAGLALAAVNGPESVVSQRTAVH  
ELEKEFAGRGREVRVLDVSHARSPLMODVLPDEFARLASVTLRPAATIPVSNVTGDL  
AGDDLADPAYWADHIRATVRPADGVRLARQVDITVVELGPDAAITLGGELDQDT

Query Match 5.0%; Score 25; DB 1; Length 76196;

Best Local Similarity 100.0%; Pred. No. 0.46; DB 1; Length 76196;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 GTGGAGGCGGACCGGACCGGACG 417

DB 44442 GTGGAGGCGGACCGGACCGGACG 44466

RESULT 44

BD269390

LOCUS BD269390 3978 bp DNA linear PAT 17-JUL-2003

DEFINITION Mitomycin biosynthetic gene cluster.

ACCESSION BD269390

VERSION BD269390.1 GI:33079158

KEYWORDS JP 2002537833-A/4.

SOURCE Streptomyces lavendulae

ORGANISM

REFERENCE

AUTHORS Sherman,D.H., Mao,Y., Varoglu,M., He,M. and Sheldon,P.C.

TITLE Mitomycin biosynthetic gene cluster

JOURNAL Patent: JP 2002537833-A 4 12-NOV-2002;

REGENTS OF THE UNIVERSITY OF MINNESOTA

COMMENT OS Streptomyces lavendulae

PN JP 2002537833-A/4

PD 12-NOV-2002

PF 10-MAR-2000 JP 2000603359

PR 12-MAR-1999 US 09/266965

PI DAVID H SHERMAN, YINGQING MAO, MUSTAFA VAROGLU, MIN HE, PAUL C PI

SHELDON

PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P17/10 PC

, C12Q1/68, G01N33/53,

PC G01N33/566//C12N9/00, (C12P17/10, C12R1:465), C12N15/00, C12N5/00

CC Mitomycin biosynthetic gene cluster

FF Key Location/Qualifiers

FT source 1. .3978

FT Location/Qualifiers

1. .3978

/organism="Streptomyces lavendulae"

/mol\_type="genomic DNA"

/db\_xref="taxon:1914"

ORIGIN

Query Match

Best Local Similarity 4.8%; Score 24; DB 6; Length 3978;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCCGACGGCGCTGCAAG 156

DB 757 CTGGCCCCCGACGGCGCTGCAAG 780

RESULT 45

AR266876

LOCUS AR266876 3978 bp DNA linear PAT 10-APR-2003

DEFINITION Sequence 19 from patent US 6495348.

ACCESSION AR266876

VERSION AR266876.1 GI:29696258

KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 3978)  
AUTHORS Sherman,D.H., Mao,Y., Varoglu,M., He,M. and Sheldon,P.  
TITLE Mitomycin biosynthetic gene cluster  
JOURNAL Patent: US 6495348-A 19 17-DEC-2002;  
FEATURES Location/Qualifiers  
source 1. .3978  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Query Match

Best Local Similarity 4.8%; Score 24; DB 6; Length 3978;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCCGACGGCGCTGCAAG 156

DB 757 CTGGCCCCCGACGGCGCTGCAAG 780

Search completed: June 16, 2004, 17:31:28

Job time : 2080.33 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 22:09:53 ; Search time 2276.33 Seconds  
(without alignments)  
9539.403 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325  
Perfect score: 501  
Sequence: 1 atgtttacacgtcgggcac.....cccggtcgagggctcgtgc 501

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.hcg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.or.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.un.\*
- 28: em.vi.\*
- 29: em.hcg.hum.\*
- 30: em.hcg.inv.\*
- 31: em.hcg.mus.\*
- 32: em.hcg.other.\*
- 33: em.hcg.mus.\*
- 34: em.hcg.pln.\*
- 35: em.hcg.rod.\*
- 36: em.hcg.mam.\*
- 37: em.hcg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	501	100.0	53784	1	AMM223012	AJ223012 Amycolato
2	501	100.0	53789	6	A69720	A69720 Sequence 3
3	501	100.0	90445	1	AF040570	AF040570 Amycolato
C 4	302.6	60.4	82746	1	AF453501	AF453501 Actinosyn
5	285	56.9	69644	1	AY179507	AY179507 Streptomy
C 6	275.4	55.0	32870	1	AF007101	AF007101 Streptomy
C 7	243.8	48.7	2614	1	ATTECR	X56928 A. teichomyc
C 8	243.8	48.7	2614	6	A21096	A21096 Synthetic D
C 9	192.2	38.4	107379	1	SHGCP1R	X86780 S. hygroscop
C 10	184.2	36.8	28732	1	AF082100	AF082100 Streptomy
C 11	184.2	36.8	77534	1	AF235504	AF235504 Streptomy
C 12	184.2	36.8	77536	6	BD235937	BD235937 Polyketid
C 13	184.2	36.8	77536	6	AR271638	AR271638 Sequence
C 14	163.8	32.7	77457	1	AF210249	AF210249 Streptomy
C 15	110.4	22.0	292200	1	SCO939129	AL939129 Streptomy
C 16	107.8	21.5	12730	1	AE005055	AE005055 Halobacte
C 17	101.2	20.2	300425	1	AP005022	AP005022 Streptomy
C 18	98.8	19.7	2520	6	AX172316	AX172316 Sequence
C 19	98.8	19.7	9984	1	ASP290449	AJ290449 Amycolato
C 20	96.2	19.2	299300	1	AP005026	AP005026 Streptomy
C 21	92.8	18.5	301675	1	AP005027	AP005027 Streptomy
C 22	91.8	18.3	1578	6	E22246	E22246 gene DNA en
C 23	90.8	18.1	13040	1	AE005769	AE005769 Caulobact
C 24	90.8	18.1	349746	1	EX572605	EX572605 Rhodopseu
C 25	89.4	17.8	309350	1	AP005215	AP005215 Corynebac
C 26	88	17.6	295150	1	SCO939125	AL939125 Streptomy
C 27	87.8	17.5	18876	6	AX803758	AX803758 Sequence
C 28	87.8	17.5	61944	6	AX803750	AX803750 Sequence
C 29	87.2	17.4	22449	1	SPSNBCDE	Y11548 S. pristinae
C 30	87.2	17.4	22449	1	SPSNBCGEN	X98690 S. pristinae
C 31	84.2	16.8	11776	1	AE004683	AE004683 Pseudomon
C 32	84	16.8	300400	1	AP005943	AP005943 Bradyrhiz
C 33	83.2	16.6	346259	1	EX640435	EX640435 Sequence
C 34	82.2	16.4	1992	8	AF052222	AF052222 Bordetell
C 35	82.2	16.4	2038	8	AF052223	AF052223 Lollum pe
C 36	81.6	16.3	299950	1	AP005963	AP005963 Bradyrhiz
C 37	81.6	16.3	346287	1	EX640450	EX640450 Bordetell
C 38	81.4	16.2	7788	6	AX803724	AX803724 Sequence
C 39	81.4	16.2	37360	6	AX803722	AX803722 Sequence
C 40	81.4	16.2	301482	1	AE016916	AE016916 Chromobac
C 41	80.8	16.1	1749	6	AR384076	AR384076 Sequence
C 42	80.6	16.1	1623	6	BD179585	BD179585 Highly th
C 43	80.6	16.1	2006	6	AX204868	AX204868 Sequence
C 44	80.4	16.0	11544	1	AF320253	AF320253 Azorarcus
C 45	80.2	16.0	1979	8	AK120964	AK120964 Oryza sat

ALIGNMENTS

RESULT 1  
AMM223012  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AMM223012 53784 bp DNA linear BCT 09-FEB-1998  
Amycolatopsis mediterranei genes encoding rifamycin polyketide  
synthases, ORFs 1 to 5.

AJ223012.1 GI:2764760  
ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.  
Amycolatopsis mediterranei  
Amycolatopsis mediterranei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.

1  
Schupp, T., Toupet, C., Engel, N. and Goff, S.  
Cloning and sequence analysis of the putative rifamycin polyketide

```

synthase gene cluster from Amycolatopsis mediterranei
Unpublished
2 (bases 1 to 53784)
Schupp, T.
Direct Submission
Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma
AG, CH 4002 Basel, Postfach K-681.344, 4002 Basel,
SWITZERLAND
FEATURES             Location/Qualifiers
     source            1..53784
                     /organism="Amycolatopsis mediterranei"
                     /mol_type="genomic DNA"
                     /strain="LBG A3136"
                     /db_xref="taxon:33910"
                     1335..15543
                     /function="CoA ligase, acylcarrier protein,
                     b-ketoacyl-ACP-synthase, acyltransferase, ketoreductase"
                     /note="ORF1
                     PKS modules 1-3"
                     /codon_start=1
                     /transl_table=11
                     /product="rifamycin polyketide synthase, type 1"
                     /protein_id="CAA11035.1"
                     /db_xref="GI:2764761"
                     /db_xref="GOA:O54666"
                     /db_xref="SPTREMBL:O54666"
                     /translation="WRTDLIKPLHVALLENATRFAGKPAFPADHRTVTYGDLEARTTR
                     LAGHLGVRGHRGDRVAICLNGRNVSTVESYFALLRAGAGVPLNPGSATAELEHPLTD
                     SGATVVTDAAQAKRLAPHELVLLTGDDVPEGNASHYDEIALSEPAEPADDDLELDE
                     PAMFYTSVTGTRKGVSTQNRCLWSVASCVPFPGSLGDDQDRLVWPLPFLHSHIA
                     CVLSATVGVSVRIADGSSADDVMLRIEASSTFLGAVPTTYHYHLVRAARGFSAPS
                     LRI GLAGAVLGLAGLSFEFTFGVPLIDAYGSTETCGAITMNPDPGARVEGSCGLAV
                     PGDVRVDPDTGLDPAVERGEVWVGNGVNLGYHNSPEATAMARDGWFRTGDLAR
                     RDDAGYFTICRIKELIRGCANLHPGEVEALRTVDCGVADAAGVPHDITLGVFVA
                     RYTPGFTGPAALEKIKREQLSAYKVPDRILEVAHIRTASGKIRGLTDEPAQLR
                     YAATHEQSHADSESAAALRUSGLDERAQCCELEDDIVRTQAADVILGQVPDGRA
                     FRDLGFTSLAIVLNRLTETHTGLWLPASAFVDFHTPPAALAAVRVAELIGITQAVAP
                     VVAADGPPIALVGNACRLPGCVASPEDLWRLVAERVDVAVSEFPDGRWDLDLSDPD
                     RERACTSVGGGFLHDAGEFDAGFGFI SPREAVAMDPOORLLLETSWEALNAGVDP
                     TALKDTGVGSLMGQGVGSAAVAPELEGFTVTCGVASSVAGRVSYVLGLGPATVV
                     DTACSSSIVAMHLAAQLRQGCSCMALAGGVVMTATPGSFVEFSQRALAPDGRCAF
                     AAADGTGWSEGVVLEIVSARERGHRIILAVIRGSANQDNGASNGITPAGLSQQ
                     RYTRALAAAGLAPSDVDVVEAHGTGTGLDPIEAQALITATYQERKQPLMGLSKSN
                     IGHAAQAGVAGVIMQVALRHETLPTLHVDPKPTLEVDMSAGAILLELTFEARAPNG
                     RRRAGSVGSGTNAHLILEEAPAEVPAAPLPVPLVVSARSSTLSQAEERLA
                     SLLEGVSLTEVAGALVSRAVLDERAVVAGSREAVTGLRALNTAGSTPGKVWV
                     PFGQGTQWAGMRELLASVPAERLAECAALAPWIDNSLVNVLREGDGRVDVLO
                     PACFAVMGLAAVWESGVKRPDAVVGHSQGETAAACVSGALSEDAKVALRSQAIA
                     AELSGGNSVALGEDVVSRLVDGEVAAVNGSSVVIAGDAHALDAILLSGEG
                     IRRVRVADYASHTRHVEDIRDTLAETLAGISAQAPAVPFTVTSEWRDAGVLDDG
                     YVYRNLNQRVFGCAATALLBOQHTVFVEVSAHPVTQPLSELGTDAIGTLRRDGL
                     RLLASMGELVRGIDVDMTAMVPAAGWDLPTYAFEHRRHYLEPAEPASAGDPLIGT
                     VVSTPGSDRLTAVQWSRAQWAVDGLVPNALVEAAILRLGLDLAGTPVVGELVDAP
                     VILPRGSEVOLII VEGPEQRRRPIEVFSREADEPWRHAGTGLTAPAAAAPPEPAA
                     GDAITVTAAGDADRGTHPALLDAAVTVVGDLLSPVTGTVSSLASGATAVTTP
                     TATGLTDPAGQVLTVESVRGTTPVFAQGTGLDPLFRVDWPEIPLPTAETADFLPYE
                     ATSAAETLSAQALWADPAETRLAVTGTCTEPGAAAIWGLVSAQSEHPGRILFADL
                     DPAPVLPAVASEGPQVRNGVASVPRLTVPTRDAPALDPEGLVTITGGTGLTGA
                     LTAHPLTVHTAGVHLVIVRGRPELQEBELTALGASVALAACDVADRAQLAEALAI
                     PAEPHTLVATHTAGVLDGVDVTELTDPRLATVRRPKVDAAALDEUTRADLAAVFLF
                     SSAAGVILGPGQGYAAANAEALDARQNSLDLPVSIANGWYATVSGMTEHLGDAD
                     LRNQRIGMSGLPADGEMALLDAIATGTLVAAKFDVAALATAKAGGVPPLPRLGL
                     APLPRAAAKTASILTRELACLAETEQAALLDVRRAAEVLHSGAESVHSRGTFO
                     AGDPSITVELNRLAACTLISPMIIFDYPKPPALADHLRAKLGSAANPAEITGT
                     AAABEPIAVAMACRFPFGVHSFEDLWRLVADGADAVTEPPADRCHDNTBELHEDPH
                     EGTTYVYHGAFLDDAAGFDAAFGFISPEALAMDPOORLLLETSWELFEEAALDPTTL
                     AGQDIIGVYAGVNSHDSYMRHRAAGVSEFRLTCGSSVLSGRVYAHFGEPAVTVDT
                     CAGSSIVLHMAQALQRGCSCMALAGGVVMTATPGSFVEFSQRGLAPDGRCKAFAD
                     ADGDTGSEGVGQVALLERLSEAQRGRGHVLAIVRGSVNSDNGSLTAPNGSQOQRV
                     IRKALAAAGLSTSDVDVAEHGTITGLDPIEAELALLATYQNRETPLMWSVKSNI
                     HTQAAAGVAGVIMQVAMRHGVLPRTHVDPRPSYVDNSAGAVELLTEARDWNSGHP
                     RRAGSVSGFGGGLVHVEVAAPITTPQEPAEFLVPVILVARSARTAGLRGQGRILA
                     AFLGDRDTRVPRDAAVALATTRAQLDHRVAVLVAASRAQLCADLAAAFSGSVYTGTPVDG

```

CDS

```

KIAVLTGQSQWAGMRELAETTFVPRDAPFAEACEAVDTHLRERPLREVVEDDALL
DOTMYTGALFAVETALFELFESWGRPGLLAGHSICELAAAHVSGVGLDADAGELVA
ARGLMQALPAGAMVAVQATEDEVAPLLDGTVCVAANGVSDVSLVLDLADAGELVA
ELAGRCRTRRLAVSHAFSTPLMEPLMDDFRAVAERLTYRAGSLPVYSTITGELAAAD
SPNYGVOVNAVRFSDATVAGCAQASTLELPGCGALAAALMGITLGGPEQSCVATL
RKNGAVPDLVLTALAEHVGRGVDMTTLVDEPATATVGTVLPTAFQHQRFWMDVDET
AAVSTPPPAEPVIDRPVQDVLBELPRESAAVWLGHRDAGSFDLDRSPKRDHGSLSAV
KLRNLRDFTGVELPSTLIFDYNPAVLADHLRAELGGERPAAPVPTKRVSDRPIAI
VGNSTRIPGGAGSPEELKWLVAERDPAVSPVDRGMDLGLVHPDPAHAGTSTYRSG
GFLHDAQFDAGLFGISPREALAMDPOORLLLETSWEALERAGVDPVLSRAGSVDGVT
GIYHVDVTRREVPEVDQGYTWTGTASSVASGRVAVYFGPEGPAVTVDTACSSLVA
MHLAAQALROGSCMALAGGATWASPDALFESRQSLGADGRCKAYABGADGTGWA
EGVGVVLERLSVARERGHVLAIVRGSANQDNGASNLTPAGSPSOQVRIRGALASA
GLAPSDVVVEGHTGTALGDPIEQALLATYQGERQPLWLSKSLNIGHTQAAAGV
VGTIKIMAMRHGVMFPAETHVDERTSODWSAGAEIVLITAREARWPRTRPRAGVSF
CAGSTNAHLIIIEEPGAEEVBEVASVPLVVSARSAGSAGQAGQALAAVLNLESAG
VAGALVSGRATLNERAVIAGSDEAQQALARGENAPGVVTGTAGKPKVWVFP
GQSQWGMGRDILDDSSPFAARIKECAALAEOWTWSLLDVLGRDADLDRSQVOP
ASFAMVGLAANWTSLGVTDPDALVGHSGQETAAACVSGALSDDAQKVALRSQAIA
ELAGRGNSVALGESEDAVARTLTPANRVEVAANVSPSVVIAGDAQALDEALALAG
DGVRRVAVDYASHTRHVEAIAETLAKTIDARVPAIPFYSTVLTWTEQAVDA
GYWYRNLNQRVRFQGSVADLAGLHTVFVEISAPHLVQPLSEISDDAVTVGSLRRDD
GGLRLILASAEELVYRGVADVMTAAVPAAGWDLPTYAFDRRHFLHEAETAEAEEM
TGTDGDELDTAEIAEFGTTFBFCADTDRDAIQAIVSEVTAVVAHADINQTSVDDT
GVADLDEFAKVTVAWLDQFDTPLDADFVFSFIAGIWGQGGQGPAGAANAVIDLA
LVEWRRARGIKATSIANGALDQIGIMDEAALAOILRRRGVTPMAPIPLAVTAMQVAG
NEKAVADMMDAAFIPIAFTSVRSPPLPADPEAKAILRAAQDDXGDTASSLADSL
RAVDPADQNRILLKLVGHASTVILGHSGABGIGPQAFQEVGFSLSAANVLRSLHA
TGLRLPATLIFDYPTEALVGYLRELLREDLDDGREDLDRVRLAANVPAFKEAG
VLDTLGLADTGTGTEPDTAETEARAPADDAEDLDAIDLSIGLVRVLAAGQTS"
15550..30759
/function="b-ketoacyl-ACP-synthase, acyltransferase,
dehydratase, ketoreductase, acylcarrier protein"
/note="ORF2
PKS modules 4-6"
/codon_start=1
/transl_table=11
/product="rifamycin polyketide synthase, type 1"
/protein_id="CAA11036.1"
/db_xref="GI:2764762"
/db_xref="GOA:O52789"
/db_xref="SPTREMBL:O52789"
/translation="MANQSWRKMSGAPNEQIVDALRASLKENVLRQENSALAAAAA
PVAIVSMACRYAGGIRGPEDPWRVYSEKADVTYGPEDRGMDEGLYHPDPNPGTTY
VREGAFLODAQDFDAGFGIISPREALAMDPOORLLLETSWEALIDPHSVGRSDI
GVYAGVQDYADPLSGEGEFNSLERALGTAGGVASGRVATLILEGPAVTVDTWCSS
SLVAITHLAQALRRGCSCMALAGSTWATPGFVGFAPQALAFDGRCKSYAAADG
LWASAGVGVLLERLSVARERGHVLAIVRGSANQDNGASNLTPAGNPAQQRVIRKA
LSAGITPDSVDVTVEGHTGTVLGPIEVQALLATYQGERQPLWLSKSLNIGHT
QAASVAGVIMQVLSIRHQLPATQHVDPATPOVDWVSAGAILLEAEGRPNGHPRR
GGISFGASGTVNAHMLISEAPEDEPTEPATQVTVPLVVSATASLAAQAGRLAEV
GVVSLADVATLVSRGMLSELDSEVFAARLAEETALGRWVDSLTWVNSLTDRGEADV
GKVVVFPQGTQWAGMRELLDSEVFAARLAEETALGRWVDSLTWVNSLTDRGEADV
LRSQALASLGRGMSVALSEEDATARLEPAWAGVEVAAVNGPSTSVIAGDAEALD
DVIDVQPASFAVMGLAAVWASLGVPEAVVYHSGQETAAACVSGALSDDAQKVAL
EALDADQGVYIRVAVDYASHTRHVEAIAETLAKTIDARVPAIPFYSTVLTWTEQ
VEDAGVLDQGYVYRNLNQRVGRGPAVELIEQGHVVFVEVSAHPVLVQPIINLVDDE
AVVGTTLRRDGGGLRLILASAEELVYRGVADVMTAAVPAAGWDLPTYAFDRRHFL
MGSTADVSLGLAGDHPHLLGAVLPQSDGLVFTSLSKSHSPWLAGHAGVGVLI
PCTVVDLALRAGDELFGVLELVEIPEALVGERGGVRVQVAVSGENFTGSRADV
SMREDDWTRHATGLIGASTSRSPDFPAWGPAGAEPIIDVNETYTLTERGYAIS
GAFQCMRAWRRSGDEVAEVALPDHRRDAGKFGHPALDLAALHTNANPDDDRSV
LTFAMNGLVHVAHSALVRVPAVPGFDALITQAADTGLVWMDSLVYRSVSAQL
ETAAGEERSLFQVDMIEVPTAETAREAEVLEAPGEAAPLELTSRVLNVAQSWLAD
AADARLVTVTVRGAEREVDTPAGAVMGLVRAAQAGVETVLTGTGSLGGIVAKHLVARH
GEPOLAVRNFASFVPRARATGEVPEAFSPGEQTVLTGTGTLGGIVKGLVAKHLVARH
GYRILVASRRGVAEDLTLELTEQATVSVVACVSDRQDVAALIAAHERPTVHLA
GLDDGVIGALNRERLAGVFAKVDVAQHLDLITRDLGLDAFVVFSSAALMGSAQGG

```



Thu Jun 17 08:25:01 2004

```

RESULT 3
AF040570          90445 bp      DNA      linear      BCT 16-APR-2001
LOCUS             Amvcolatopsis mediterranei rifamycin biosynthetic gene cluster.
DEFINITION
ACCESSION         AF040570 AF040571
VERSION           AF040570.2 GI:12331604
KEYWORDS
SOURCE
ORGANISM          Amvcolatopsis mediterranei
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
1 (bases 1 to 90445)
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R., Yu, T.W.,
Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X., Hutchinson, C.R. and
Floss, H.G.
Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
the molecular analysis of the rif biosynthetic gene cluster of
Amycolatopsis mediterranei S699
Chem. Biol. 5 (2), 69-79 (1998)
9512878
2 (bases 1 to 90445)
Kim, C.G., Yu, T.W., Fryhle, C.B., Handa, S. and Floss, H.G.
3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
formation of the precursor of mC7N units in rifamycin and related
antibiotics
J. Biol. Chem. 273 (11), 6030-6040 (1998)
98165773
9497318
3 (bases 1 to 90445)
Yu, T.W., Muller, R., Muller, M., Zhang, X., Draeger, G., Kim, C.G.,
Leistner, E. and Floss, H.G.
Mutational analysis and reconstituted expression of the
biosynthetic genes involved in the formation of
3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin
biosynthesis in amycolatopsis mediterranei S699
J. Biol. Chem. 276 (16), 12546-12555 (2001)
21201076
11278540
4 (bases 1 to 90445)
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R.,
Hutchinson, C.R. and Floss, H.G.
Direct Submission
Submitted (30-DEC-1997) Dept. of Chemistry, University of
Washington, Box 351700, Seattle, WA 98195-1700, USA
5 (bases 1 to 90445)
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R.,
Hutchinson, C.R. and Floss, H.G.
Direct Submission
Submitted (19-JAN-2001) Dept. of Chemistry, University of
Washington, Box 351700, Seattle, WA 98195-1700, USA
REMARK
Sequence update by submitter
On or before Jan 22, 2001 this sequence version replaced
gi:2792340, gi:2832263.
FEATURES
source
1..90445
/organism="Amycolatopsis mediterranei"
/mol_type="genomic DNA"
/strain="S699"
/db_xref="taxon:33910"
423..1451
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC01706.1"
/db_xref="GI:2792310"
/translation="MPTGVGESRYVKTLRQLEVALIGAGLRIHLRLAWLGAGAAV
RVYSDGRRELAEEFGAKAGLEALGADVDICTTASHHEIALTAIAAGVGV
CKPLAASAEAEAEIIVTAERAGRVLYAAHDVFAAPYARLHELIVASRLGELGRF
SFSAHPRWTHASAGSGLITDQLHGDLAHVFGDVRVHACYQGDIAIPFEG
AVATGTAIVLTHASGALSQVSVRTATPRPVRVAPHVSGTGSVSDSWPQEVRYD
GGAGNFAYGGPSVFTMEFEATAFAGGPEPRIGAKDALAAVRRIHAAAEASWTGRAV
ELPVRGAA"
CDS
1448..2215
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC01707.1"
/db_xref="GI:2792311"
/translation="MKVAIISSSTPQAYAGALRGLPDVVEVAAASWDAPEPVRQAAEAG
ARVCEYPPAPKATKDLKAMIDAGRLTFPASPACHGEAFVAVRKGIDAGGIGELTTLV
GSVATSDVGLGAAAPYLLDLADAVLGGBPAQCYAQTNIIVISGRIGESAALVTVYR
SGVASDFCDRRHSGATGLPVTFFIGDQSGVQYDAGPOLLGGERPELGGEDLEALMLKD
FLGAGDGPDPQAALRTFRIIAAYESAHTGQPVDL"
2353..2619
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC01708.1"
/db_xref="GI:2792312"
/translation="WQDSVLGDPRKQVEAAMSVTQVKIALSGGPAELAPAHRSVDSG
ALENTLKIRHGAGYEHFVHGGEFQAVDGCDDIAVFWMSHRTKIAE"
2652..3842
/note="ORF0; similar to Streptomyces Sp. Chop"
/codon_start=1
/transl_table=11
/product="cytochrome P450 monooxygenase"
/protein_id="AAC01709.1"
/db_xref="GI:2792313"
/translation="MTDAISFEVPMWDRDKDPDPVAVDSLRERPLAKMVYPDGHVGM
TVSSYELVRELSDLRFHSCEVHPVTHQGVPTHTPLIPGMFIHMPPEHTRVYRK
LLTGFTVRRASRLIPRAEVAEQIENVRAKGAADVMDFAKPLVRLMLGSLVGLP
YEERDYPVATLLHDAEADPAEAAAEVAGKFDEVIERRQRQDDLISSLVTE
LYQEEURNIVTLLFAGYETTGALATGVFALLHTDQLAAAEKELDAAIEELLR
YITNQYTHIYTALEDVKLEBELIKGDTVTYVSLPAANDPKAPCAELDIERTSG
HVAFGFIHQICLQNLRIELRAGFTALLRPAFELRPAVPADEVPLKLGVSFVSKKL
PVSW"
3883..18090
/gene="rifa"
3883..18090
/gene="rifa"
/note="Rifa"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAC01710.1"
/db_xref="GI:2792314"
/translation="MRTDLIKLHVALLLENATRFAGKPAFADDERHTVYGDLEARTER
LAGHLAGLVGHGDRVAICLGNRVSTVSYFAILRAGAVGVPFLNPGSATALEHPLTD
SGATVVTDAQAARLARLAPHEVLLVTGDDVPEGAHSYDELAISEPAEPADLELDE
PAMFVTSGTTRPKGVVSTORNCISVASCYVFPGLSDQDRVLPFLPHFSULSLA
CVLSATVGVASRIADGSSADDVMRLTEAESSTFLAGVPTTYHHLVRAARGFSAPS
LRIGLAGVALGRSEFEFTFGVPLIDAYGSTETCGAITMNPDPGARVSGSLAV
PGVDVVPDPTGLDVPAGEGEVWVSGPNVLMGLHNSPENTAAAMDGWERTGDLAR
RDDAGYFTICRIKELIIRGGANTHPGEAVLRITVDGVADAAGVGVPHDILGRVPA
YVIPGTGFDPALIEKREQLSAYKVPDRILEVAHTPTASGKIRGLILDFEPAQLR
YAATHEQSRHADESVAALRARLGLDERAQCELLEDLVRTQAADVGLQFVDPDGRA
FRDLGFTSLAIVELNRLTEHTGWLNPASAVFDHTPAALAAVFAELLLGTQVAEP
VVAADPGEPITAVGMKCLPGCVASPEDWFLVAERVDVASEFPQDRMDLSDILDPD
RERAGTSYVGGGFLHDAGFEFAGFGLSPREAVAMPQORLLLETSEALDEGLVQ
IALKTDITGVFSGMLQGGYSGAVAPLEGVTVMTATPGSFVFEFRQRALADGRCKAF
DTACSSILAMHLAAQALQGCESMALGEGVTVNATPGSVNODGASNGILTPNGLSQ
AAAADGTGWSRGVGVVLERLSVARERGHRIILAVRGSADVQDGSNGILTPNGLSQ
RVIRALAAAGLAPSDVDVVEAGHTTGLDPIEAQALLATYQGERKQPLWGLKSN
IGHAQAAGVAGVINKVQALRHETILPPTLHYDKTFLVDWSAGATLTLTEARWPRNG
RPRRAGVSSFGVSGTNALILIEAPAEFVVAAPPELVVPLVVSARSTESLSQAERLA
SILLEGDSLITEVAGALVSRVLDERAVVAGSREEAVTGLRALNTASGTPGVVWV
PFGQCTQWAGMGRELIAESPFAERIAECAALAPWIDSLVDVLRGSGDGRVQVLO
PACFAVMYGLAAWESVGRPDVAVVHGQGEIAAASCVSALDEADKAVLARSOAIA
AELSGRGNASVALGEDDVLSLVGVGVAAVNGSPSVIAGDAHALDAILSGEG
ITVRVAVDYASHTRHVEDIIRLTLETLAGLSAQAPVFPYSTVTSWVRDAGVLDDG
YMYRNLNRQVRFGAAATLDEQHTTVFVEVSHAPVIVQPLSELTDGATLREDDGL
RELLASMGELFVRGIDVDMTAVPAAGVMDLPTLYAFEPHHRHWELEPAEPASAGDPLIGT
VYVTSQSDRLTVAQWSRAQPAWGVMDVLPNALVEAAIIRLGLDGLTPVVLVDIAP
VYVPRGSEVOLIIVGEQGEORRRPTEVFSREADEPWTIRHAGHTLAPAAAVPEPAAA
GUAOTDVTVAGLRDADRYGHPALLDAAVRTVTVGDDLLPSVWTVGSVLSLASCATAVTVP

```







gene  
CDS  
/transl\_table=11  
/product="RNA polymerase sigma factor"  
/protein\_id="AA54109.1"  
/db\_xref="GI:21449377"  
/translation="MWIGISADGRRRVRWVWVTEAFWTETDEDIKLAATLERDAEA  
ANADSTRAYLKAIGRVRLLTADQVELARRVEAGLYAEHLDDLPAAAGSRPVELR  
RVVRDQARADHLEARNLRLVSIKRYTGKGMALDVIQENGLVRAVEKEDYAKG  
EKFSVTWIRQAITRALADQSRTIRLPHVVEVVTKLARLERDLHNRGREATAE  
VALELDVTEKVELRQAHARTFVSLDQTIGAEADTSGIDLIADENAVAPLEAVSSALL  
SADLRVAVLTLPFREAGVVRMYGFDPGLHTLDEVRVYGVTRERIRQIESRTWAKL  
QHPARSGLEGYLD"  
complement (6501. .9518)  
/gene="asm30"  
complement (6501. .9518)  
/gene="asm30"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450"  
/protein\_id="AA54108.1"  
/db\_xref="GI:21449376"  
/translation="MVATGTRIPGPKPLVGNLLDVLTSDDTDVDFLDRCHREHG  
IVALTGAGQOVFASSELVAGVHPALQVDRDFAGDGLFTARGDEPNW  
GKAHLLMPAFGFTAMRDHFPMDLTAQMLVWRPGRDHRIDVADDMTRLTJDTIA  
LCAGFARFYRDRAPFVDVVRSLVEAGRAERLPGVQFVLGNRQYRDDIATM  
NRADGIVAAARALPAGEREDDILLERMLTADPVGTGERLSARNVRYQLATFLIAGHET  
TSGLLSPAVHRLAHPLEVRKAKADVDGLDVPDEFEDLARLDYLGQVRETLRLHP  
TAPAFALDPBAEPALGHAIGAGEPVLWMLPTLHRDPVAVWRDPDVFDPERFAPRME  
IPACNMPFGHGACIGRPFALQEAFLVIALVLFQFDLADPDHRLTIKQTLTKP  
LIVRARPRADPRGATVETVYVQVQVPAHRTGTHLVYSGNSGSGEGLARTIAG  
DGAARWATVPADDDAVRALPAGSPVTVSSSNGAPPDNAHFRVRLIQDGPDLSC  
VDVLGCGNLSATYQRYPTLIDEAAGARRLRERGAITDARADFQDWRWYEP  
LWPLLSACGVEGEIGRFRVSVESDAADGLDGLASAVLENRELVRGPDGSKHLE  
LRLPDGTSYRTGTVLSYLPQNHPDLVRRVRAERCPPRRAELAAATGATLLERLFPSCAVD  
RVDELTPCYDLSAPAGVAVRLAERCPPRRAELAAATGATLLERLFPSCAVD  
LALAEHLPAPRTRLYSISNAEQRABVALTVSVGTGTVSGYLSRVPGRDVAVGIAS  
PPESFRPADNTVPVLLIAGTGIAFPGRFLRAALGGEGPALLFGCGEPDLDDL  
YAEFFAALGDWLEVDRAYSRHPDGEVHVQRLWRRDRVRVELVDAGARVILCGDATR  
VGPAVEVIGRIGFGAGWLDALRAGRYATDVF"  
9664. .10245  
/gene="asm29"  
9664. .10245  
/gene="asm29"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="AA54107.1"  
/db\_xref="GI:21449375"  
/translation="MSRAVRREQLVAALRVKRDGIAAATRAICAEADMPHGAFHY  
CFRSKQELYTALLTGLNTSLDAWPAISDPLPAVDNVRTMLAFAWTFQADPERELV  
LGLVLTQAQREPELPAWQREYVAKVAHLDRLAEEAGFDYSVDRLAELVVS  
LDGVASSWFSHRDGLAKASLDQPAVMPAALT"  
10393. .10857  
/gene="asm28"  
10393. .10857  
/gene="asm28"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AA54106.1"  
/db\_xref="GI:21449374"  
/translation="MTDTTTRHTPTVQNAALAVAGVFLVIGIAGVFGLTNTYDGLF  
AGHSDAALGVFKVFNHLNHLVLAFGAAGIAAGLSKAPRGARFLICGGAIYAVLWL  
YGVVVDHSTANFVPLNAADNLHLGLAVGMIGLGAALGRNAAHYRAARSDD"  
complement (10912. .11937)  
/gene="asm27"  
complement (10912. .11937)  
/gene="asm27"

gene  
CDS  
/transl\_table=11  
/product="bicyclomycin resistant protein"  
/protein\_id="AA54113.1"  
/db\_xref="GI:21449381"  
/translation="MAYKADSAGVGMPLALLI VALAGTII EIDMSVSPSPDIAREIG  
VGGPAVQVLTWNFLGCLGCHGPLSDRGRPVLGGSAVNLGALCALAPTIG  
LLLPVQVIGASAAVAVFVITIGDXYRGAALRMVGLYNASALMTCPALGGAI  
NAATGNGNLYAVCATLALLVFLFELPETRTARTAITVRSADYRLILSGRAYL  
NRVSTGELLFAAYLVFAASPLYFAAFGLGTGFALHLLVVAASFAPSAFTGIIP  
LLGGPAGAVRGLGAVVGLGFLLLGGDPVAVTGVSVALFCVGFVCPVYVFDMSV  
FPFELAGASSATMGLRALLVTLTLLALAGALSTGPFAPAAVMAGAAVAVLLSAPR  
"  
complement (3450. .4064)  
/gene="asm34"  
complement (3450. .4064)  
/gene="asm34"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="AA54112.1"  
/db\_xref="GI:21449380"  
/translation="MARNPERRAALLDAVEVLAGEGARGLTFRANDQAGVPAGTA  
SNYPAGRAEILTAGERVYQRLDQAPLGETPDGPRDSRVTELMHALVDRVSAPPTG  
FLALLELRLEARNPELREVLTVARADLRNABHHAASGLPGDGTVALLWLNWL  
ILRLTLPDLITDHRDLVTLVDRLLAGQPEAGSPFDESSEK"  
4151. .4723  
/gene="asm33"  
4151. .4723  
/gene="asm33"  
/codon\_start=1  
/transl\_table=11  
/product="dihydrofolate reductase"  
/protein\_id="AA54111.1"  
/db\_xref="GI:21449379"  
/translation="MRKLVIYFAASIDGVIAGPEGHDFFYPTAEPNVEHLRAEYPTT  
PTHVPLVGMPLDTPNRRYDVTVMGRGVSQPALDAGVSPYAHLRQYVVSRLPDADP  
VALVRAQPVGLVRKEDGLDMLCGADLAGQLPEIDELVVKTYVAVAGGCTPL  
FVAGFAPRAFPRVDRAFDHGGVVTYTPA"  
complement (4854. .5057)  
/gene="asm32"  
complement (4854. .5057)  
/gene="asm32"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AA54110.1"  
/db\_xref="GI:21449378"  
/translation="MWITQONARIIPARRSSGRSNARRCAVRWPRQPSRATRAPGPRS  
"

gene  
CDS  
/transl\_table=11  
/product="RNA polymerase sigma factor"  
/protein\_id="AA54109.1"  
/db\_xref="GI:21449377"  
/translation="MWIGISADGRRRVRWVWVTEAFWTETDEDIKLAATLERDAEA  
ANADSTRAYLKAIGRVRLLTADQVELARRVEAGLYAEHLDDLPAAAGSRPVELR  
RVVRDQARADHLEARNLRLVSIKRYTGKGMALDVIQENGLVRAVEKEDYAKG  
EKFSVTWIRQAITRALADQSRTIRLPHVVEVVTKLARLERDLHNRGREATAE  
VALELDVTEKVELRQAHARTFVSLDQTIGAEADTSGIDLIADENAVAPLEAVSSALL  
SADLRVAVLTLPFREAGVVRMYGFDPGLHTLDEVRVYGVTRERIRQIESRTWAKL  
QHPARSGLEGYLD"  
complement (6501. .9518)  
/gene="asm30"  
complement (6501. .9518)  
/gene="asm30"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450"  
/protein\_id="AA54108.1"  
/db\_xref="GI:21449376"  
/translation="MVATGTRIPGPKPLVGNLLDVLTSDDTDVDFLDRCHREHG  
IVALTGAGQOVFASSELVAGVHPALQVDRDFAGDGLFTARGDEPNW  
GKAHLLMPAFGFTAMRDHFPMDLTAQMLVWRPGRDHRIDVADDMTRLTJDTIA  
LCAGFARFYRDRAPFVDVVRSLVEAGRAERLPGVQFVLGNRQYRDDIATM  
NRADGIVAAARALPAGEREDDILLERMLTADPVGTGERLSARNVRYQLATFLIAGHET  
TSGLLSPAVHRLAHPLEVRKAKADVDGLDVPDEFEDLARLDYLGQVRETLRLHP  
TAPAFALDPBAEPALGHAIGAGEPVLWMLPTLHRDPVAVWRDPDVFDPERFAPRME  
IPACNMPFGHGACIGRPFALQEAFLVIALVLFQFDLADPDHRLTIKQTLTKP  
LIVRARPRADPRGATVETVYVQVQVPAHRTGTHLVYSGNSGSGEGLARTIAG  
DGAARWATVPADDDAVRALPAGSPVTVSSSNGAPPDNAHFRVRLIQDGPDLSC  
VDVLGCGNLSATYQRYPTLIDEAAGARRLRERGAITDARADFQDWRWYEP  
LWPLLSACGVEGEIGRFRVSVESDAADGLDGLASAVLENRELVRGPDGSKHLE  
LRLPDGTSYRTGTVLSYLPQNHPDLVRRVRAERCPPRRAELAAATGATLLERLFPSCAVD  
RVDELTPCYDLSAPAGVAVRLAERCPPRRAELAAATGATLLERLFPSCAVD  
LALAEHLPAPRTRLYSISNAEQRABVALTVSVGTGTVSGYLSRVPGRDVAVGIAS  
PPESFRPADNTVPVLLIAGTGIAFPGRFLRAALGGEGPALLFGCGEPDLDDL  
YAEFFAALGDWLEVDRAYSRHPDGEVHVQRLWRRDRVRVELVDAGARVILCGDATR  
VGPAVEVIGRIGFGAGWLDALRAGRYATDVF"  
9664. .10245  
/gene="asm29"  
9664. .10245  
/gene="asm29"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="AA54107.1"  
/db\_xref="GI:21449375"  
/translation="MSRAVRREQLVAALRVKRDGIAAATRAICAEADMPHGAFHY  
CFRSKQELYTALLTGLNTSLDAWPAISDPLPAVDNVRTMLAFAWTFQADPERELV  
LGLVLTQAQREPELPAWQREYVAKVAHLDRLAEEAGFDYSVDRLAELVVS  
LDGVASSWFSHRDGLAKASLDQPAVMPAALT"  
10393. .10857  
/gene="asm28"  
10393. .10857  
/gene="asm28"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AA54106.1"  
/db\_xref="GI:21449374"  
/translation="MTDTTTRHTPTVQNAALAVAGVFLVIGIAGVFGLTNTYDGLF  
AGHSDAALGVFKVFNHLNHLVLAFGAAGIAAGLSKAPRGARFLICGGAIYAVLWL  
YGVVVDHSTANFVPLNAADNLHLGLAVGMIGLGAALGRNAAHYRAARSDD"  
complement (10912. .11937)  
/gene="asm27"  
complement (10912. .11937)  
/gene="asm27"

```

/contig_start=1
/transl_table=11
/product="glucose 1-dehydrogenase"
/protein_id="AAM54105.1"
/db_xref="GI:21449373"
/transl_table=11
/transl_table="MPRRAOVVVVVVGVASSGIGRVTLRFAAAGARVVCAARNTAL
DGLVVRGAGRAVAVTHADIAEAARAVADVAVVERFGRVDTWNAAGIGVYGRVED
TPAGEFDRVVRVYLGRVHGAALPALRRAGGVGLIGVASVLGLSAPLQAPYARSK
Query Match 60.4%; Score 302.6; DB 1; Length 82746;
Best Local Similarity 75.2%; Pred. No. 4.7e-27;
Matches 377; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 1 ATGTTCTACAGTCGGGACGACGCGGCGGCCCAAGGGCGTCTGTCCACGACGCGCAAC 60
Db 36558 ATGCTCTACACCTCGGACACACCGCGGCCCAAGGGCGTGTGTCTACCCAGCGCAAC 36499
QY 61 TGCCTCTGGTCGCTCGCTTCTGCTAGTCGCTTCCCGCGGTGTCTGGACCGACCGG 120
Db 36498 TGCCTGTGTCGCTGGCGGCTGCTAGCGCGCGCTGCTGCTGCTGACCCCGACCGG 36439
QY 121 GTGCTCTGGCGGCTCCCGCTGTCTCCACAGCGCTTTCGACATCGCTCGCTGTCGCGC 180
Db 36438 GTGCTGTGGCGGCTCGCGCTGTCTCCACAGCTGTTCGACATCGCGGCTGTGCGCAC 36379
QY 181 ACGTGTGTCGGGCGACGCTCGGATCGCGACGCGAGCTCCGCGGACGAGCTGATCGG 240
Db 36378 ACGCGCTGCGGCGCACACCGCGGTGTTCGACGCGGCTCTCCCGCGAGGAGCTGCTG 36319
QY 241 CTGATCGAGCGGAGAGCTCGACCTCTCTGCGCGGCGTCCGACACCTACACCACTG 300
Db 36318 GCGCTGACAGGAGGCGCTCCACGCTGCTCGCGGCGTCCGACCACTACCACTG 36259
QY 301 GTGCGGCGCGCGCGCGCGGCTTCTCGCGCGGAGCTTCGCGATCGGCTGCGCGG 360
Db 36258 GTGCGCGGCGCGGTCGAGCGCGGCTTCTCGCGCGGAGCTTCGCGATCGGCTGCG 36199
QY 361 GCGCGGCTCTCGGCGCGCGGCTCGGAGCGAGCTTCGAGAGACCTTCGCGGTCGCG 420
Db 36198 GCGCGGATCACACGCGCGGCTCGGCGCGGCTTCGAGCGAGCTTCGCGGTCGCG 36139
QY 421 ATCGAGCGCTACGCGACGACGCGAGCTTCGCGGCGGATCACCATGACCGCGGACG 480
Db 36138 ATCGAGCGCTACGCGACGACGCGAGCTTCGCGGCGGATCACCATGACCGCGG 36079
QY 481 GCCCGCTGCGAGGCTCGTGC 501
Db 36078 CCCAGGTCGAGGCTCCAGC 36058

RESULT 5
AY179507 69644 bp DNA linear BCT 21-FEB-2003
LOCUS
DEFINITION Streptomyces hygroscopicus strain NRRL 3602 geldanamycin
biosynthesis gene cluster, partial sequence.
ACCESSION AY179507
VERSION AY179507.1 GI:28192593
KEYWORDS
SOURCE
ORGANISM Streptomyces hygroscopicus
Streptomyces hygroscopicus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 69644)
Rascher, A., Hu, Z., Viswanathan, N., Schirmer, A., Reid, R.,
Nierman, W.C., Lewis, M. and Hutchinson, C.R.
Cloning and characterization of a gene cluster for geldanamycin
production in Streptomyces hygroscopicus NRRL 3602
FEMS Microbiol. Lett. 218 (2), 223-230 (2003)
22474162
PUBMED 12586396
2 (bases 1 to 69644)
Rascher, A., Hu, Z., Viswanathan, N., Schirmer, A., Reid, R.,
Nierman, W.C., Lewis, M. and Hutchinson, C.R.
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
```

```

TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Kosan Biosciences, 3832 Bay Center Place,
Hayward, CA 94545, USA
FEATURES
source
1. 69644
/organism="Streptomyces hygroscopicus"
/mol_type="genomic DNA"
/strain="NRRL 3602"
/db_xref="taxon:1912"
complement(1..684)
/notes="ORF1c"
/codon_start=1
/transl_table=11
/product="similar to predicted RhtB family transporters"
/protein_id="AAO06909.1"
/db_xref="GI:28192594"
/transl_table="MSFGRSDSGGHHVSTESVLTFAAMSLVIVIPGSLVFWGR
ALAHGRRTALVIGVICTYALVAVANGALGVESVALFMVGLAGAYLVYLVG
QAFHRRMRVANMDAPADERGDLRSILDILVGTNPKAGIVFFAVALPQFVDHSVG
HVPVQMVGLVLPITICLVTDLTWGLGSAARSFWARSRRLSMWVGAGGLAMIGLV
TVAATGRAD"
complement(771..1673)
/notes="similar to S. coelicolor SC3C8.01 (SC6382); ORF17"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="AAO06910.1"
/db_xref="GI:28192595"
/transl_table="MKRTMGRHRTIQPRTTTLATRAVLAAGVLVPTIASAGSAHA
TPQATICTORPELADKLSDEINSALGSAATTAISHDRTNVTCULDRHFDSDAS
TPQVLTSLMDAQDNRAITOEKDHAMITESDNDATTALWKQLGADKINGFLQ
AGMTWTTDSEHGVLQITANDKEKLLQVLTNPNVLSDDSRAYILKULATRVPSQ
RWGTPAGASDAOVHVKNGLERATNGRVHSLGAFTEGDDHDTITVLSQDNATMDGG
IANLEGTAHVHENLAPVSSAQ"
complement(1945..2736)
/notes="similar to S. coelicolor SCF1.09 (SC0267); ORF18"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="AAO06911.1"
/db_xref="GI:28192596"
/transl_table="MTINVPVAGGSLDVRVCGDMGPTLVFAHYWGGSDTWNVGIGH
PPGATVRFQDQGMGTGRALPGVHLDDLDLVRVIERVSGPPVLVHSGMGKAGQ
LVAAATPAGLVGLVAPPPQPPAMVTEYRQGLSHAYDSAEVTHALDHLVITATPL
PEAVRTAERDSLAAGAEARQEWPLGIARDITDAAREIEVPTVLAARENDVVEPHI
LREHLLPHIPLATLTTPVEDAGHLLPAEPAGVARALGDFLTAIQ"
2890..3774
/notes="AraC family; similar to S. coelicolor SC5A7.14
(SC6644) and SCF1.08c (SC0266); ORF19"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06912.1"
/db_xref="GI:28192597"
/transl_table="MENGSAVRQLRYLPVAGSAGVEVLDFALRSMQDTRRRTQPO
RPDFHVALVSGTSGHEADFENYGLGEGGAVIRPGMVHRWSDIDACDPLILRPPQ
FLSGFTAGATAPACWHLDRQLSLALLAAHLGHEHSTAVHTPLASVLLSLAA
LITRALGTPPSVGPSPASRPTVEFVRAAVERFTDWHVADYALGVDVRLT
RTTRATGTGAKTFLDQRLLEAKELLATDLPVSGCARRLGRDVGNGFTTFRQAG
LPFAAARAAYTAGAGG"
complement(3824..4585)
/notes="Tetr family; similar to S. coelicolor 2SCD46.08
(SC4194); ORF20"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06913.1"
/db_xref="GI:28192598"
/transl_table="MLSVNTVTPSHLNAPRLTRGKYIPGABIYQGSATYDAVDC
PGLRERKQRTAAISDAISLFLVGENQVSAOVAEAAEYKRTLPAYFTKEDLV
VHRLADHETESARVVRAPPTAPITALREHLEGLERDPTGLNDQPEVRLHBMV
LDAPSVARMAGFKTGAERALTAEKRETDFTELTARLAQVIVAHWALAQDNAERL
AYGEPADERYAGAVADMEDIAFALLENGHLLPFRQ"
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Kosan Biosciences, 3832 Bay Center Place,
Hayward, CA 94545, USA
FEATURES
source
1. 69644
/organism="Streptomyces hygroscopicus"
/mol_type="genomic DNA"
/strain="NRRL 3602"
/db_xref="taxon:1912"
complement(1..684)
/notes="ORF1c"
/codon_start=1
/transl_table=11
/product="similar to predicted RhtB family transporters"
/protein_id="AAO06909.1"
/db_xref="GI:28192594"
/transl_table="MSFGRSDSGGHHVSTESVLTFAAMSLVIVIPGSLVFWGR
ALAHGRRTALVIGVICTYALVAVANGALGVESVALFMVGLAGAYLVYLVG
QAFHRRMRVANMDAPADERGDLRSILDILVGTNPKAGIVFFAVALPQFVDHSVG
HVPVQMVGLVLPITICLVTDLTWGLGSAARSFWARSRRLSMWVGAGGLAMIGLV
TVAATGRAD"
complement(771..1673)
/notes="similar to S. coelicolor SC3C8.01 (SC6382); ORF17"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="AAO06910.1"
/db_xref="GI:28192595"
/transl_table="MKRTMGRHRTIQPRTTTLATRAVLAAGVLVPTIASAGSAHA
TPQATICTORPELADKLSDEINSALGSAATTAISHDRTNVTCULDRHFDSDAS
TPQVLTSLMDAQDNRAITOEKDHAMITESDNDATTALWKQLGADKINGFLQ
AGMTWTTDSEHGVLQITANDKEKLLQVLTNPNVLSDDSRAYILKULATRVPSQ
RWGTPAGASDAOVHVKNGLERATNGRVHSLGAFTEGDDHDTITVLSQDNATMDGG
IANLEGTAHVHENLAPVSSAQ"
complement(1945..2736)
/notes="similar to S. coelicolor SCF1.09 (SC0267); ORF18"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="AAO06911.1"
/db_xref="GI:28192596"
/transl_table="MTINVPVAGGSLDVRVCGDMGPTLVFAHYWGGSDTWNVGIGH
PPGATVRFQDQGMGTGRALPGVHLDDLDLVRVIERVSGPPVLVHSGMGKAGQ
LVAAATPAGLVGLVAPPPQPPAMVTEYRQGLSHAYDSAEVTHALDHLVITATPL
PEAVRTAERDSLAAGAEARQEWPLGIARDITDAAREIEVPTVLAARENDVVEPHI
LREHLLPHIPLATLTTPVEDAGHLLPAEPAGVARALGDFLTAIQ"
2890..3774
/notes="AraC family; similar to S. coelicolor SC5A7.14
(SC6644) and SCF1.08c (SC0266); ORF19"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06912.1"
/db_xref="GI:28192597"
/transl_table="MENGSAVRQLRYLPVAGSAGVEVLDFALRSMQDTRRRTQPO
RPDFHVALVSGTSGHEADFENYGLGEGGAVIRPGMVHRWSDIDACDPLILRPPQ
FLSGFTAGATAPACWHLDRQLSLALLAAHLGHEHSTAVHTPLASVLLSLAA
LITRALGTPPSVGPSPASRPTVEFVRAAVERFTDWHVADYALGVDVRLT
RTTRATGTGAKTFLDQRLLEAKELLATDLPVSGCARRLGRDVGNGFTTFRQAG
LPFAAARAAYTAGAGG"
complement(3824..4585)
/notes="Tetr family; similar to S. coelicolor 2SCD46.08
(SC4194); ORF20"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06913.1"
/db_xref="GI:28192598"
/transl_table="MLSVNTVTPSHLNAPRLTRGKYIPGABIYQGSATYDAVDC
PGLRERKQRTAAISDAISLFLVGENQVSAOVAEAAEYKRTLPAYFTKEDLV
VHRLADHETESARVVRAPPTAPITALREHLEGLERDPTGLNDQPEVRLHBMV
LDAPSVARMAGFKTGAERALTAEKRETDFTELTARLAQVIVAHWALAQDNAERL
AYGEPADERYAGAVADMEDIAFALLENGHLLPFRQ"
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Kosan Biosciences, 3832 Bay Center Place,
Hayward, CA 94545, USA
FEATURES
source
1. 69644
/organism="Streptomyces hygroscopicus"
/mol_type="genomic DNA"
/strain="NRRL 3602"
/db_xref="taxon:1912"
complement(1..684)
/notes="ORF1c"
/codon_start=1
/transl_table=11
/product="similar to predicted RhtB family transporters"
/protein_id="AAO06909.1"
/db_xref="GI:28192594"
/transl_table="MSFGRSDSGGHHVSTESVLTFAAMSLVIVIPGSLVFWGR
ALAHGRRTALVIGVICTYALVAVANGALGVESVALFMVGLAGAYLVYLVG
QAFHRRMRVANMDAPADERGDLRSILDILVGTNPKAGIVFFAVALPQFVDHSVG
HVPVQMVGLVLPITICLVTDLTWGLGSAARSFWARSRRLSMWVGAGGLAMIGLV
TVAATGRAD"
complement(771..1673)
/notes="similar to S. coelicolor SC3C8.01 (SC6382); ORF17"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="AAO06910.1"
/db_xref="GI:28192595"
/transl_table="MKRTMGRHRTIQPRTTTLATRAVLAAGVLVPTIASAGSAHA
TPQATICTORPELADKLSDEINSALGSAATTAISHDRTNVTCULDRHFDSDAS
TPQVLTSLMDAQDNRAITOEKDHAMITESDNDATTALWKQLGADKINGFLQ
AGMTWTTDSEHGVLQITANDKEKLLQVLTNPNVLSDDSRAYILKULATRVPSQ
RWGTPAGASDAOVHVKNGLERATNGRVHSLGAFTEGDDHDTITVLSQDNATMDGG
IANLEGTAHVHENLAPVSSAQ"
complement(1945..2736)
/notes="similar to S. coelicolor SCF1.09 (SC0267); ORF18"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="AAO06911.1"
/db_xref="GI:28192596"
/transl_table="MTINVPVAGGSLDVRVCGDMGPTLVFAHYWGGSDTWNVGIGH
PPGATVRFQDQGMGTGRALPGVHLDDLDLVRVIERVSGPPVLVHSGMGKAGQ
LVAAATPAGLVGLVAPPPQPPAMVTEYRQGLSHAYDSAEVTHALDHLVITATPL
PEAVRTAERDSLAAGAEARQEWPLGIARDITDAAREIEVPTVLAARENDVVEPHI
LREHLLPHIPLATLTTPVEDAGHLLPAEPAGVARALGDFLTAIQ"
2890..3774
/notes="AraC family; similar to S. coelicolor SC5A7.14
(SC6644) and SCF1.08c (SC0266); ORF19"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06912.1"
/db_xref="GI:28192597"
/transl_table="MENGSAVRQLRYLPVAGSAGVEVLDFALRSMQDTRRRTQPO
RPDFHVALVSGTSGHEADFENYGLGEGGAVIRPGMVHRWSDIDACDPLILRPPQ
FLSGFTAGATAPACWHLDRQLSLALLAAHLGHEHSTAVHTPLASVLLSLAA
LITRALGTPPSVGPSPASRPTVEFVRAAVERFTDWHVADYALGVDVRLT
RTTRATGTGAKTFLDQRLLEAKELLATDLPVSGCARRLGRDVGNGFTTFRQAG
LPFAAARAAYTAGAGG"
complement(3824..4585)
/notes="Tetr family; similar to S. coelicolor 2SCD46.08
(SC4194); ORF20"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06913.1"
/db_xref="GI:28192598"
/transl_table="MLSVNTVTPSHLNAPRLTRGKYIPGABIYQGSATYDAVDC
PGLRERKQRTAAISDAISLFLVGENQVSAOVAEAAEYKRTLPAYFTKEDLV
VHRLADHETESARVVRAPPTAPITALREHLEGLERDPTGLNDQPEVRLHBMV
LDAPSVARMAGFKTGAERALTAEKRETDFTELTARLAQVIVAHWALAQDNAERL
AYGEPADERYAGAVADMEDIAFALLENGHLLPFRQ"
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Kosan Biosciences, 3832 Bay Center Place,
Hayward, CA 94545, USA
FEATURES
source
1. 69644
/organism="Streptomyces hygroscopicus"
/mol_type="genomic DNA"
/strain="NRRL 3602"
/db_xref="taxon:1912"
complement(1..684)
/notes="ORF1c"
/codon_start=1
/transl_table=11
/product="similar to predicted RhtB family transporters"
/protein_id="AAO06909.1"
/db_xref="GI:28192594"
/transl_table="MSFGRSDSGGHHVSTESVLTFAAMSLVIVIPGSLVFWGR
ALAHGRRTALVIGVICTYALVAVANGALGVESVALFMVGLAGAYLVYLVG
QAFHRRMRVANMDAPADERGDLRSILDILVGTNPKAGIVFFAVALPQFVDHSVG
HVPVQMVGLVLPITICLVTDLTWGLGSAARSFWARSRRLSMWVGAGGLAMIGLV
TVAATGRAD"
complement(771..1673)
/notes="similar to S. coelicolor SC3C8.01 (SC6382); ORF17"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="AAO06910.1"
/db_xref="GI:28192595"
/transl_table="MKRTMGRHRTIQPRTTTLATRAVLAAGVLVPTIASAGSAHA
TPQATICTORPELADKLSDEINSALGSAATTAISHDRTNVTCULDRHFDSDAS
TPQVLTSLMDAQDNRAITOEKDHAMITESDNDATTALWKQLGADKINGFLQ
AGMTWTTDSEHGVLQITANDKEKLLQVLTNPNVLSDDSRAYILKULATRVPSQ
RWGTPAGASDAOVHVKNGLERATNGRVHSLGAFTEGDDHDTITVLSQDNATMDGG
IANLEGTAHVHENLAPVSSAQ"
complement(1945..2736)
/notes="similar to S. coelicolor SCF1.09 (SC0267); ORF18"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="AAO06911.1"
/db_xref="GI:28192596"
/transl_table="MTINVPVAGGSLDVRVCGDMGPTLVFAHYWGGSDTWNVGIGH
PPGATVRFQDQGMGTGRALPGVHLDDLDLVRVIERVSGPPVLVHSGMGKAGQ
LVAAATPAGLVGLVAPPPQPPAMVTEYRQGLSHAYDSAEVTHALDHLVITATPL
PEAVRTAERDSLAAGAEARQEWPLGIARDITDAAREIEVPTVLAARENDVVEPHI
LREHLLPHIPLATLTTPVEDAGHLLPAEPAGVARALGDFLTAIQ"
2890..3774
/notes="AraC family; similar to S. coelicolor SC5A7.14
(SC6644) and SCF1.08c (SC0266); ORF19"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06912.1"
/db_xref="GI:28192597"
/transl_table="MENGSAVRQLRYLPVAGSAGVEVLDFALRSMQDTRRRTQPO
RPDFHVALVSGTSGHEADFENYGLGEGGAVIRPGMVHRWSDIDACDPLILRPPQ
FLSGFTAGATAPACWHLDRQLSLALLAAHLGHEHSTAVHTPLASVLLSLAA
LITRALGTPPSVGPSPASRPTVEFVRAAVERFTDWHVADYALGVDVRLT
RTTRATGTGAKTFLDQRLLEAKELLATDLPVSGCARRLGRDVGNGFTTFRQAG
LPFAAARAAYTAGAGG"
complement(3824..4585)
/notes="Tetr family; similar to S. coelicolor 2SCD46.08
(SC4194); ORF20"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06913.1"
/db_xref="GI:28192598"
/transl_table="MLSVNTVTPSHLNAPRLTRGKYIPGABIYQGSATYDAVDC
PGLRERKQRTAAISDAISLFLVGENQVSAOVAEAAEYKRTLPAYFTKEDLV
VHRLADHETESARVVRAPPTAPITALREHLEGLERDPTGLNDQPEVRLHBMV
LDAPSVARMAGFKTGAERALTAEKRETDFTELTARLAQVIVAHWALAQDNAERL
AYGEPADERYAGAVADMEDIAFALLENGHLLPFRQ"
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Kosan Biosciences, 3832 Bay Center Place,
Hayward, CA 94545, USA
FEATURES
source
1. 69644
/organism="Streptomyces hygroscopicus"
/mol_type="genomic DNA"
/strain="NRRL 3602"
/db_xref="taxon:1912"
complement(1..684)
/notes="ORF1c"
/codon_start=1
/transl_table=11
/product="similar to predicted RhtB family transporters"
/protein_id="AAO06909.1"
/db_xref="GI:28192594"
/transl_table="MSFGRSDSGGHHVSTESVLTFAAMSLVIVIPGSLVFWGR
ALAHGRRTALVIGVICTYALVAVANGALGVESVALFMVGLAGAYLVYLVG
QAFHRRMRVANMDAPADERGDLRSILDILVGTNPKAGIVFFAVALPQFVDHSVG
HVPVQMVGLVLPITICLVTDLTWGLGSAARSFWARSRRLSMWVGAGGLAMIGLV
TVAATGRAD"
complement(771..1673)
/notes="similar to S. coelicolor SC3C8.01 (SC6382); ORF17"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="AAO06910.1"
/db_xref="GI:28192595"
/transl_table="MKRTMGRHRTIQPRTTTLATRAVLAAGVLVPTIASAGSAHA
TPQATICTORPELADKLSDEINSALGSAATTAISHDRTNVTCULDRHFDSDAS
TPQVLTSLMDAQDNRAITOEKDHAMITESDNDATTALWKQLGADKINGFLQ
AGMTWTTDSEHGVLQITANDKEKLLQVLTNPNVLSDDSRAYILKULATRVPSQ
RWGTPAGASDAOVHVKNGLERATNGRVHSLGAFTEGDDHDTITVLSQDNATMDGG
IANLEGTAHVHENLAPVSSAQ"
complement(1945..2736)
/notes="similar to S. coelicolor SCF1.09 (SC0267); ORF18"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="AAO06911.1"
/db_xref="GI:28192596"
/transl_table="MTINVPVAGGSLDVRVCGDMGPTLVFAHYWGGSDTWNVGIGH
PPGATVRFQDQGMGTGRALPGVHLDDLDLVRVIERVSGPPVLVHSGMGKAGQ
LVAAATPAGLVGLVAPPPQPPAMVTEYRQGLSHAYDSAEVTHALDHLVITATPL
PEAVRTAERDSLAAGAEARQEWPLGIARDITDAAREIEVPTVLAARENDVVEPHI
LREHLLPHIPLATLTTPVEDAGHLLPAEPAGVARALGDFLTAIQ"
2890..3774
/notes="AraC family; similar to S. coelicolor SC5A7.14
(SC6644) and SCF1.08c (SC0266); ORF19"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06912.1"
/db_xref="GI:28192597"
/transl_table="MENGSAVRQLRYLPVAGSAGVEVLDFALRSMQDTRRRTQPO
RPDFHVALVSGTSGHEADFENYGLGEGGAVIRPGMVHRWSDIDACDPLILRPPQ
FLSGFTAGATAPACWHLDRQLSLALLAAHLGHEHSTAVHTPLASVLLSLAA
LITRALGTPPSVGPSPASRPTVEFVRAAVERFTDWHVADYALGVDVRLT
RTTRATGTGAKTFLDQRLLEAKELLATDLPVSGCARRLGRDVGNGFTTFRQAG
LPFAAARAAYTAGAGG"
complement(3824..4585)
/notes="Tetr family; similar to S. coelicolor 2SCD46.08
(SC4194); ORF20"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06913.1"
/db_xref="GI:28192598"
/transl_table="MLSVNTVTPSHLNAPRLTRGKYIPGABIYQGSATYDAVDC
PGLRERKQRTAAISDAISLFLVGENQVSAOVAEAAEYKRTLPAYFTKEDLV
VHRLADHETESARVVRAPPTAPITALREHLEGLERDPTGLNDQPEVRLHBMV
LDAPSVARMAGFKTGAERALTAEKRETDFTELTARLAQVIVAHWALAQDNAERL
AYGEPADERYAGAVADMEDIAFALLENGHLLPFRQ"
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Kosan Biosciences, 3832 Bay Center Place,
Hayward, CA 94545, USA
FEATURES
source
1. 69644
/organism="Streptomyces hygroscopicus"
/mol_type="genomic DNA"
/strain="NRRL 3602"
/db_xref="taxon:1912"
complement(1..684)
/notes="ORF1c"
/codon_start=1
/transl_table=11
/product="similar to predicted RhtB family transporters"
/protein_id="AAO06909.1"
/db_xref="GI:28192594"
/transl_table="MSFGRSDSGGHHVSTESVLTFAAMSLVIVIPGSLVFWGR
ALAHGRRTALVIGVICTYALVAVANGALGVESVALFMVGLAGAYLVYLVG
QAFHRRMRVANMDAPADERGDLRSILDILVGTNPKAGIVFFAVALPQFVDHSVG
HVPVQMVGLVLPITICLVTDLTWGLGSAARSFWARSRRLSMWVGAGGLAMIGLV
TVAATGRAD"
complement(771..1673)
/notes="similar to S. coelicolor SC3C8.01 (SC6382); ORF17"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="AAO06910.1"
/db_xref="GI:28192595"
/transl_table="MKRTMGRHRTIQPRTTTLATRAVLAAGVLVPTIASAGSAHA
TPQATICTORPELADKLSDEINSALGSAATTAISHDRTNVTCULDRHFDSDAS
TPQVLTSLMDAQDNRAITOEKDHAMITESDNDATTALWKQLGADKINGFLQ
AGMTWTTDSEHGVLQITANDKEKLLQVLTNPNVLSDDSRAYILKULATRVPSQ
RWGTPAGASDAOVHVKNGLERATNGRVHSLGAFTEGDDHDTITVLSQDNATMDGG
IANLEGTAHVHENLAPVSSAQ"
complement(1945..2736)
/notes="similar to S. coelicolor SCF1.09 (SC0267); ORF18"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="AAO06911.1"
/db_xref="GI:28192596"
/transl_table="MTINVPVAGGSLDVRVCGDMGPTLVFAHYWGGSDTWNVGIGH
PPGATVRFQDQGMGTGRALPGVHLDDLDLVRVIERVSGPPVLVHSGMGKAGQ
LVAAATPAGLVGLVAPPPQPPAMVTEYRQGLSHAYDSAEVTHALDHLVITATPL
PEAVRTAERDSLAAGAEARQEWPLGIARDITDAAREIEVPTVLAARENDVVEPHI
LREHLLPHIPLATLTTPVEDAGHLLPAEPAGVARALGDFLTAIQ"
2890..3774
/notes="AraC family; similar to S. coelicolor SC5A7.14
(SC6644) and SCF1.08c (SC0266); ORF19"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06912.1"
/db_xref="GI:28192597"
/transl_table="MENGSAVRQLRYLPVAGSAGVEVLDFALRSMQDTRRRTQPO
RPDFHVALVSGTSGHEADFENYGLGEGGAVIRPGMVHRWSDIDACDPLILRPPQ
FLSGFTAGATAPACWHLDRQLSLALLAAHLGHEHSTAVHTPLASVLLSLAA
LITRALGTPPSVGPSPASRPTVEFVRAAVERFTDWHVADYALGVDVRLT
RTTRATGTGAKTFLDQRLLEAKELLATDLPVSGCARRLGRDVGNGFTTFRQAG
LPFAAARAAYTAGAGG"
complement(3824..4585)
/notes="Tetr family; similar to S. coelicolor 2SCD46.08
(SC4194); ORF20"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06913.1"
/db_xref="GI:28192598"
/transl_table="MLSVNTVTPSHLNAPRLTRGKYIPGABIYQGSATYDAVDC
PGLRERKQRTAAISDAISLFLVGENQVSAOVAEAAEYKRTLPAYFTKEDLV
VHRLADHETESARVVRAPPTAPITALREHLEGLERDPTGLNDQPEVRLHBMV
LDAPSVARMAGFKTGAERALTAEKRETDFTELTARLAQVIVAHWALAQDNAERL
AYGEPADERYAGAVADMEDIAFALLENGHLLPFRQ"
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Kosan Biosciences, 3832 Bay Center Place,
Hayward, CA 94545, USA
FEATURES
source
1. 69644
/organism="Streptomyces hygroscopicus"
/mol_type="genomic DNA"
/strain="NRRL 3602"
/db_xref="taxon:1912"
complement(1..684)
/notes="ORF1c"
/codon_start=1
/transl_table=11
/product="similar to predicted RhtB family transporters"
/protein_id="AAO06909.1"
/db_xref="GI:28192594"
/transl_table="MSFGRSDSGGHHVSTESVLTFAAMSLVIVIPGSLVFWGR
ALAHGRRTALVIGVICTYALVAVANGALGVESVALFMVGLAGAYLVYLVG
QAFHRRMRVANMDAPADERGDLRSILDILVGTNPKAGIVFFAVALPQFVDHSVG
HVPVQMVGLVLPITICLVTDLTWGLGSAARSFWARSRRLSMWVGAGGLAMIGLV
TVAATGRAD"
complement(771..1673)
/notes="similar to S. coelicolor SC3C8.01 (SC6382); ORF17"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="AAO06910.1"
/db_xref="GI:28192595"
/transl_table="MKRTMGRHRTIQPRTTTLATRAVLAAGVLVPTIASAGSAHA
TPQATICTORPELADKLSDEINSALGSAATTAISHDRTNVTCULDRHFDSDAS
TPQVLTSLMDAQDNRAITOEKDHAMITESDNDATTALWKQLGADKINGFLQ
AGMTWTTDSEHGVLQITANDKEKLLQVLTNPNVLSDDSRAYILKULATRVPSQ
RWGTPAGASDAOVHVKNGLERATNGRVHSLGAFTEGDDHDTITVLSQDNATMDGG
IANLEGTAHVHENLAPVSSAQ"
complement(1945..2736)
/notes="similar to S. coelicolor SCF1.09 (SC0267); ORF18"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="AAO06911.1"
/db_xref="GI:28192596"
/transl_table="MTINVPVAGGSLDVRVCGDMGPTLVFAHYWGGSDTWNVGIGH
PPGATVRFQDQGMGTGRALPGVHLDDLDLVRVIERVSGPPVLVHSGMGKAGQ
LVAAATPAGLVGLVAPPPQPPAMVTEYRQGLSHAYDSAEVTHALDHLVITATPL
PEAVRTAERDSLAAGAEARQEWPLGIARDITDAAREIEVPTVLAARENDVVEPHI
LREHLLPHIPLATLTTPVEDAGHLLPAEPAGVARALGDFLTAIQ"
2890..3774
/notes="AraC family; similar to S. coelicolor SC5A7.14
(SC6644) and SCF1.08c (SC0266); ORF19"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06912.1"
/db_xref="GI:28192597"
/transl_table="MENGSAVRQLRYLPVAGSAGVEVLDFALRSMQDTRRRTQPO
RPDFHVALVSGTSGHEADFENYGLGEGGAVIRPGMVHRWSDIDACDPLILRPPQ
FLSGFTAGATAPACWHLDRQLSLALLAAHLGHEHSTAVHTPLASVLLSLAA
LITRALGTPPSVGPSPASRPTVEFVRAAVERFTDWHVADYALGVDVRLT
RTTRATGTGAKTFLDQRLLEAKELLATDLPVSGCARRLGRDVGNGFTTFRQAG
LPFAAARAAYTAGAGG"
complement(3824..4585)
/notes="Tetr family; similar to S. coelicolor 2SCD46.08
(SC4194); ORF20"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06913.1"
/db_xref="GI:28192598"
/transl_table="MLSVNTVTPSHLNAPRLTRGKYIPGABIYQGSATYDAVDC
PGLRERKQRTAAISDAISLFLVGENQVSAOVAEAAEYKRTLPAYFTKEDLV
VHRLADHETESARVVRAPPTAPITALREHLEGLERDPTGLNDQPEVRLHBMV
LDAPSVARMAGFKTGAERALTAEKRETDFTELTARLAQVIVAHWALAQDNAERL
AYGEPADERYAGAVADMEDIAFALLENGHLLPFRQ"
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Kosan Biosciences, 3832 Bay Center Place,
Hayward, CA 94545, USA
FEATURES
source
1. 69644
/organism="Streptomyces hygroscopicus"
/mol_type="genomic DNA"
/strain="NRRL 3602"
/db_xref="taxon:1912"
complement(1..684)
/notes="ORF1c"
/codon_start=1
/transl_table=11
/product="similar to predicted RhtB family transporters"
/protein_id="AAO06909.1"
/db_xref="GI:28192594"
/transl_table="MSFGRSDSGGHHVSTESVLTFAAMSLVIVIPGSLVFWGR
ALAHGRRTALVIGVICTYALVAVANGALGVESVALFMVGLAGAYLVYLVG
QAFHRRMRVANMDAPADERGDLRSILDILVGTNPKAGIVFFAVALPQFVDHSVG
HVPVQMVGLVLPITICLVTDLTWGLGSAARSFWARSRRLSMWVGAGGLAMIGLV
TVAATGRAD"
complement(771..1673)
/notes="similar to S. coelicolor SC3C8.01 (SC6382); ORF17"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="AAO06910.1"
/db_xref="GI:28192595"
/transl_table="MKRTMGRHRTIQPRTTTLATRAVLAAGVLVPTIASAGSAHA
TPQATICTORPELADKLSDEINSALGSAATTAISHDRTNVTCULDRHFDSDAS
TPQVLTSLMDAQDNRAITOEKDHAMITESDNDATTALWKQLGADKINGFLQ
AGMTWTTDSEHGVLQITANDKEKLLQVLTNPNVLSDDSRAYILKULATRVPSQ
RWGTPAGASDAOVHVKNGLERATNGRVHSLGAFTEGDDHDTITVLSQDNATMDGG
IANLEGTAHVHENLAPVSSAQ"
complement(1945..2736)
/notes="similar to S. coelicolor SCF1.09 (SC0267); ORF18"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="AAO06911.1"
/db_xref="GI:28192596"
/transl_table="MTINVPVAGGSLDVRVCGDMGPTLVFAHYWGGSDTWNVGIGH
PPGATVRFQDQGMGTGRALPGVHLDDLDLVRVIERVSGPPVLVHSGMGKAGQ
LVAAATPAGLVGLVAPPPQPPAMVTEYRQGLSHAYDSAEVTHALDHLVITATPL
PEAVRTAERDSLAAGAEARQEWPLGIARDITDAAREIEVPTVLAARENDVVEPHI
LREHLLPHIPLATLTTPVEDAGHLLPAEPAGVARALGDFLTAIQ"
2890..3774
/notes="AraC family; similar to S. coelicolor SC5A7.14
(SC6644) and SCF1.08c (SC0266); ORF19"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06912.1"
/db_xref="GI:28192597"
/transl_table="MENGSAVRQLRYLPVAGSAGVEVLDFALRSMQDTRRRTQPO
RPDFHVALVSGTSGHEADFENYGLGEGGAVIRPGMVHRWSDIDACDPLILRPPQ
FLSGFTAGATAPACWHLDRQLSLALLAAHLGHEHSTAVHTPLASVLLSLAA
LITRALGTPPSVGPSPASRPTVEFVRAAVERFTDWHVADYALGVDVRLT
RTTRATGTGAKTFLDQRLLEAKELLATDLPVSGCARRLGRDVGNGFTTFRQAG
LPFAAARAAYTAGAGG"
complement(3824..4585)
/notes="Tetr family; similar to S. coelicolor 2SCD46.08
(SC4194); ORF20"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAO06913.1"
/db_xref="GI:28192598"
/transl_table="MLSVNTVTPSHLNAPRLTRGKYIPGABIYQGSATYDAVDC
PGLRERKQRTAAISDAISLFLVGENQVSAOVAEAAEYKRTLPAYFTKEDLV
VHRLADHETESARVVRAPPTAPITALREHLEGLERDPTGLNDQPEVRLHBMV
LDAPSVARMAGFKTGAERALTAEKRETDFTELTARLAQVIVAHWALAQDNAERL
AYGEPADERYAGAVADMEDIAFALLENGHLLPFRQ"
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Kosan Biosciences, 3832 Bay Center Place,
Hayward, CA 94545, USA
FEATURES
source
1. 69644
/organism="Streptomyces hygroscopicus"
/mol_type="genomic DNA"
/strain="NRRL 3602"
/db_xref="taxon:1912"
complement(1..684)
/notes="ORF1c"
/codon_start=1
/transl_table=11
/product="similar to predicted RhtB family transporters"
/protein_id="AAO06909.1"
/db_xref="GI:281
```

```

gene
4569. .6065
/gene="gdmL"
CDS
4569. .6065
/gene="gdmL"
/notes="putative FAD-dependent monooxygenase"
/codon_start=1
/transl_table=11
/product="GdmL"
/protein_id="AA006914.1"
/db_xref="GI:28192599"
/translation="MLTSTEVVAVAGAGATGMLAYELALAGVETLVLEKLQRIQQ
VKGTQTPRTAEILEGRLELPEMLRLAIARADPDGSGFALPFLVDCAPWTEHPFIF
IPOWERLEVEERAEAGAVLRGTAVGVAPDDGGVVVADGLRAHAYLVACDGGH
STVKLLGLPPFGRAGTHAVLADIRLSAVSLVPRQMGIMTWTHARGYWMVLP
GDRYPTFGHAQDARTPTWTHIEIAALQAVGPETIGAVDNSSRFSFATQQL
EHTGRVLPAQAGAHHPPLGAGLNLGVQDANLNGKLAALVQNRHLAGNGLDSHA
ERHFAAQLVHSHAAQVLAISNPSEDAALRDIPTDLRLPQNRHLAGNGLSLR
YDLPGHPITGERIPADLVETGTTRLSLTFSGSHAVLJDLAGAVPADLPPLPRVDL
VRATCADDMAAALLRPDGVVCWATDTSAAAGDTLLAALTGDLARVP"
complement(6107. .6577)
/gene="gdmX"
complement(6107. .6577)
/gene="gdmX"
/notes="similar to Jadx and MmyY"
/codon_start=1
/transl_table=11
/product="GdmX"
/protein_id="AA006915.1"
/db_xref="GI:28192600"
/translation="MTATADYDSLARLRLEALKEALRMIRGWRALDRKOWTWE
CWAEADVLEFGPEMKIKGKAVRAKVEVESPVASMQHILNMHVDVVEGDRATGIGYM
WFVAVTAPGTSSTSPYSNGPFWDFEFRPGDWLLVRQRLGVWMTDGEDTLRAFA"
7208. .27736
/gene="gdmAI"
7208. .27736
/gene="gdmAI"
/notes="polyketide synthase; modules 0-3"
/codon_start=1
/transl_table=11
/product="GdmAI"
/protein_id="AA006916.1"
/db_xref="GI:28192601"
/translation="MLRTDLIRPVPELIRANADRFGDKPACDGHRTVSHAELERRTR
RLAGHLARLGDURAMICLGNREVMVESYFGLRVANGVAVPVNPSTDAELSLLA
DSGARLVLDVAHQDRLEQPELVVVGDPPLKGFIAPELPDTPQALAAD
DLGDELITWLYTSGTGLPKGVLSIQBNCLMSLAAACVPTVGLTAEDRVLWELPLH
SLSHIVCLLAATVAGASTRIVDGVSADVLDALEERSTFIAGVTPLHHLIEAERER
DFATPELRIALVGVAVATDILVRFESAFGLVDVAYGSTETCGAIAVNNPTGRVREG
SCGLFVPLITVLPDPTVDVDPAGREGFWVGNPTMAGYHNOPEATASALRDGWR
TGLDRDEAGCTVTGRIKELVIRAGENIHPEGEVAVLRTVPGVADAADVVKPHAVL
GEVPAFVPGPDGFDPSALILATCRERLSYFKVPEEIVEIARVPTSLAGKTRHVLLE
LPARLRAGDQVDSLLRLWVPOSALPDAPAGTGTWALADADALGLAVGLRAAGVDA
RVNGEPGEFVADSVAGLVAGSVADLGDGGAAPDVVVVTPPVAGLPDETGADEAGV
TVGERDLARLGAWLADRLAGTTFVVTATGAVATGAEDADEPELSAALWGVVRSLL
QAAYFGRLLTVLDLGDAGDAGDGREALLRAVQGHQQAATIRGGVLLVPLRL
TRISFPAGPAPTLADAGLAVITGGDTTGRGTALRHLYATYAGRNILLISANGLPPE
AAALRTLELRDQGVMSVCDPGRDALTDSLVDAQAPPTAAVHIEPSPERSIDTS
LRAMTHLEERTGRGAAPALFVVVTSNAGVLSGSPGRDRAAQQFGEALVRRREALGLG
LALAWGLPFGHGHTAVAGAVPLPEALAFDAAALTAGOGPLVLLRPSITGLPGGEPVP
AVLRHLVDAPGSFADSPAVAFVPRFLAASESGRQRMALAVREHAATLGLASAD
PVEAQAFSAFTGTSLTVALRNRLNAATGARLAATVVDHPTPAGLARHIVREITGR
RGVQAPVARGVSDPVAIVAMGCHLPGEVATPEDLWRLVADGRDIALAGFVREDRWDL
AGLFDSDPAVGVSVRGGFTLDAGGFDAAFRGLSPREALAMPQORLLILTAWEYF
ENAGIDPGSLGHTDGVFSVGMVHDYGDAGTAABGLBGLGVSGSAGSVSVGRVAYAL
GLTGAFTVDTACSSSLVHLHVAQVATGECSSALAGVAVMSRPTSFIFBSRQGL
APDGRCKSFAGEDATNWSBEGVLLRLSDARENHEVLAVRGTRAVNQDGSNGL
TAPNGPSQERVIRQALNAGLTADVDAVEAHGTGSLGDTPEAQALLATYQERPED
QPLWLSLKNIGHAQAAGAGVTKVQAMRHGVLPKTLHADDEPTSKVDWTSAGVSL
LSEAPWPEPTGPBAGTSSFGVSCTNHVVLEQAPLEAAPTQASDAGAPGLVATG
GVVPWLSAKTPAALRAQERLVSLESGSDANFVDVGNLSLTTAALEHRAVLIATD
AEGGWATARALESERPDLLVTGOTDGTGVFFIPGQGAQVGMGAQLNLTSPVFAA
RLRECADALAPDWSLIDVITGTPDAPSLERVDVQVATFVAVVSLAALWQSVGIHP
DAVIGHSQEIATAACVAGHLTUTNAKIVTLRSQRTIAHLAGHGMSLATPADTIDL

```

```

Query Match 56.9%; Score 285; DB 1; Length 69644;
Best Local Similarity 73.1%; Pred. No. 5.4e-25;
Matches 366; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 1 ATGTTCTACAGCTCGGCGCAGCACCGGCGCGCCCAAGGCGCTGCTGTCCACGACGCGCAAC 60
DB 7715 ATGCTCTACACCTCGGCGCACCA CGGCGCTCGCGAAGAGCGTGTCTCCACACAGCGGAC 7774

QY 61 TGCCTCTGCTCGCTCGCTTCTGCTACGTCGCGCTTCCCGGGTGTTCGACACGAGACCG 120
DB 7775 TGCTGTGCTTCCCTCGCGCGCTGCTACTGTCGCGGTGACCGGCGCTGACCGCGAGACCGC 7834

QY 121 GTGCTCTGCGCGCTCGCGCTGTTTCCACAGCCTTTTCGCACATCGCCTCGCTCTCGCGCC 180
DB 7835 GTGCTGTGCGCGCTCGCGCTGTTTCCACAGCCTTTTCGCACATCGCTGTCTGCTCGCGGCC 7894

QY 181 ACCCTGTGTCGGGGCCAGCGTCCGGATCGCCGACGCGCAGCTCCGCGCAGCAGCGTATCGG 240
DB 7895 ACCCGCTGTGCGGGCCAGCACCCGGATGCTGACGCGGGTGTTCGACGGCGGATGCTGCGAC 7954

QY 241 CTGATCAGAGCGGAGAGCTGACCTTCTTGGCGCGGCTGCGGACACACTACACACCTG 300
DB 7955 GCACTGCGCGAGGAGCGCTGACCTTCTATCGCGGAGTGCAGCGCTCTTACACACCTG 8014

QY 301 GTGCGGGCGCGCGGACGCGCGGTTTCTCGCGCGGAGCTTCGGATCGCGCTCGCGCGG 360
DB 8015 ATCAGAGCGCGCGCGGCGCGACTTTCGCCACGCGCGAGCTGCGGATCGCGCTCGTGGGC 8074

QY 361 GGCGCGGCTCTCGCGCGCGGCTCGAAGCGAGTTCGAAGAGACCTTCGGGGGTCCCGCTG 420
DB 8075 GGGCGGTGCGCACCGCGGACCTGTCAGTCTGTCGAGTCCGCTTCGAGTGCCTC 8134

QY 421 ATCAGCGCTTACGCGCAGCACCGAGACTGTCGCGGGCGATCACCATGAAACCCCGGACCGC 480
DB 8135 GTCAGCGCTTACGCGCTCACCGGAGACCTGTGCGCGATCGCGGTCAAATGCGCAACCGCGC 8194

QY 481 GCCCGCTCGAGGGGTCTGTGC 501
DB 8195 CCGCGGTGCGAGGGGTCTGTGC 8215

```

```

TNWKGKLTAAHNSPNATVIAGDTDALHOLHTHYTDQGTARLIIPVDYASHTGHVDTI
KNQODVLGDTILEPCTIPLWSTVTGWIPEPNTVGDYSYWRNLRQVQPEHTLHTLAD
QGYRTIELSPHPVLTAIOETLEANDTNTTIVTCTLRDDDTPLRLTNLAHTTN
GFPNNPILFTGTQTRIFLPTYPFQHHYWLPRNTSAGDSVAGVGTGTHGVLAVV
SVPDTCGVLTLQSLVATHPLADHVAHVGVTLPLFTAMARLAIARAGETDITPHELV
IQGPMTLBEDGALHVQLVGGVEDGRRVRIYSRYPDAAQEQEWLEHASKGTLA TOPDGS
AEGGMENGPWPFPQGVLEPTALDDFDYDLAQAGYEYGPAPGLKVKRKGDFVFAEA
LPEQDQVAGRFGIHPHALLDAALHASNFCVPPAPGOTLLPFVWNGVRLLAAGATAVRV
RAPATCTDFTTSLFSTGSPVASVSLVIRALSPQLAAAGSGRGSADALTLDWT
EHTALTGTEVNATLGDATDVAHDALLAGEDRPAVVADTAANAAGDTGLPARAR
DLAARALDIVORWVPELADVRLVLTREGANSVHDIAETVTPDAAALINGLVRSAQSE
HGRALIDVTDVDSREALPEAAVEQVALRRGLLWVPRVSPQGLVPAHEHWTL
DYSEKLSNLLVLRPEATAPLTCQVRIEVAAGCNFEDVLVAGVAGQGLGGE
GAGVTEVGVGVEGLAVGRVGLFPRSGPLAIADARTVAPIPEGSYATAAGVPVA
YLTAIXGLRDLGTQVGETVLHAAAGVGMAAVOLARHFGATVYATAHSGKHVLTIA
LVGPEGLASSRDVGFASAPFALDVNLSTLGEVVDASLGLTGGRFVEMGKNDIRD
PASVAAHVGVGQAFDLGDGADPDRIRRELLAELVELFEAGREPLPIRHWDTQAPT
AFWMSQGRHTGKILVLTFRALPDGVTGLITGGTGLGATIAHVVTHHARGALLIS
RQEPDAPGATDITTELTELGATVRIATACDTRDQLAALADIIPAAHPLTAVIHTAGA
LDGVLTLTPDRLDVFRKVDATHLMDLTDQLAAFIYSAAGTIGNAQANY
AAANFLDAFAOWRHARHPATSLAWGLMSDTSTLTMDTADVTRRTRAGVLGMDNA
EALRVDTGLRSGRPAALVAKIDLTALRAPDAELSLRLGLARPARTARTATAAAGG
LSCQLAGLSPAGOREFILLNVRAEAVILAHAGPEAIEPTVAKEMGFDSLTAVLEN
RLNATATGLRLPDLFDPHTFALLTELFTHELGGGPAPAAAAPVTVRAADEPLAVVA
MSCRLPGVTDPLDGLWNLLEERDGIADPDRDGDLEALFDADDPDRSGTYSYLRGFG
LEDAAGFDADFGISPREALAMPQORLFLCAWVEVFERAGMDPTTVGGGDIYGFAGV
INODYVRSQGAPELDEGLVMTGSAVSASRVAYVGLRGPATVVDTCSSSLVAMH
WAYOALRQECMSMALAGGATVMGRPSAFVFSRQGLAPDGLCKAFAGADGTTFSBG
VGVILLERSDARRNGHEVLAVIRGTAVNQDGSNGLTAPNGSQQRVIRQALANAGL
SATVIDAVEAHGTALGDPIEAQALLATYQDRPGDPEPVLGSLKNTSKNTHHAAV
SSYIKMVLAMRHGTLPRSYADEPTPEVDSQAGVSLLTTEARPEWPTSHERRAGISF
GIGTWNHLLLEAQPSQETEEPAAPKADGMDTPGLVATGGSVPWLSAKTPALRAQ
AQRLDHLSESGVTDRLDLCWSLATTTRTLHDHRAILLTDEGGDATAALTALATQPH

```

```

RESULT 6
AF007101
LOCUS
DEFINITION
  32870 bp DNA linear BCT 13-JAN-1998
  Streptomyces hygroscopicus putative pteridine-dependent
  dioxigenase, PKS modules 1, 2, 3 and 4, and putative regulatory
  protein genes, complete cds and putative hydroxylase gene, partial
  cds.
ACCESSION
  AF007101
VERSION
  AF007101.1 GI:2624946
KEYWORDS
  Streptomyces hygroscopicus
SOURCE
  Streptomyces hygroscopicus
  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
  Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
  1 (bases 1 to 32870)
  Ruan, X., Staasi, D., Lax, S.A. and Katz, L.
  A second type-I PKS gene cluster isolated from Streptomyces
  hygroscopicus ATCC 29253, a rapamycin-producing strain
  Gene 203 (1), 1-9 (1997)
JOURNAL
  MEDLINE
  PUBMED
  9426000
REFERENCE
  2 (bases 1 to 32870)
  Ruan, X. and Lax, S.
  Direct Submission
  Submitted (05-JUN-1997) 47PAP9A, Abbott Laboratories, 100 Abbott
  Park Road, Abbott Park, IL 60064, USA
  Location/Qualifiers
    1..32870
    /organism="Streptomyces hygroscopicus"
    /mol_type="genomic DNA"
    /strain="ATCC 29253"
    /db_xref="taxon:1912"
    complement(108..1130)
    /note="orf5; similar to RapK"
    /codon_start=1
    /transl_table=11
    /product="putative pteridine-dependent dioxigenase"
    /protein_id="AAC38060.1"
    /db_xref="GI:2624947"
    /translation="MNPSSVILNGLTSVFENGARVVPVGVGNILGVNVASVCEYPT
    LDHGVPLEINMAPTAEPAEVAVVTVDAESPHGERDGIWVAHDGEYFCAGRPVPTGR
    YTEATRAAYVMFELLSEFGYSSVFRMNEIGDINRDAECMEYRPFCEGSAEAPQ
    CRLEDFDPAATGIGSGGGGIAFYLLACRSGGHVHIENPRQVPAHHPKRYCPAPRF
    ARATYLPRAADGGGVGVSGTASVLGHETAHGDLVKQCRLENIELVLSGNGLA
    AHGISAGHGLTALRNIKYVRRSDVPVAREICREAFSPDADIVLYTVDCRSLLIVE
    IEGVVM"
    1454..9625
    /note="orf1; loading domain"
    /codon_start=1
    /transl_table=11
    /product="PKS module 1"
    /protein_id="AAC38061.1"
    /db_xref="GI:2624948"
    /translation="WLRRELRPLPELLKAAHARGSOTAFADSRGVSVAQLAERT
    RLAGHLQRLPGRDRAIYIGNCVETTESIALARASAIAGVFINPHSSTAEALYLD
    DSAARVITDTRILPQVRRLQARPHLVTVVYDGEAGVAFDAAETEPQAPDRL
    GLDIAWLVYSGTTPKPGVLSQSCLSVAAS SAGVLGLSETDRVILFVPLFSL
    AHIVLVAVATGATRTAEAPDAEELLHRAEFTFLAGVTVHYLLDAARDGDL
    SLPLRVCLSAIATSTLRAEFQFTGVLPLDCVGSFTCTQLITANWNGTQAGSC
    GLVPGLRVIVDETHESVETGEGEIHWGSPSLMVGNHYNQPEATEALRGWYRTG
    DLRGDDIDGHYHTITGLSELIRSGENLHPTVEVEQVLKRVPGVTDAAVVGSAHPTLGE
    VPAFLVPAADGDFDPEYFAACREHAYKVPPELYEIGSPIRTSGSKIKKHALAQSP
    ARLVAVSGSPESFLREWTAPVPEAGVDPGQATSAHRAGSVTAGLEAFGSAVDC
    PDGSLJEYVALRTGLTDRSGRAVLTHGADHPQAQAVWGLVSMRADAPDAFVLY
    DADGGEALATSAVLASQCPALVRGGLVPLRLARAGSGORVSEPDGTVLVRGAD
    QEQAREAGHLVAAYGIRHVRLLARTAVPDLVARLTTELGSATVQTVRAPGRASIAEA
    FQTGPPVTVAVLRAGPETLPTTLTGVEVLEPLQDALDELTRAHPALFVSLAPID
    NPGHAFVQAVNRHAAGLPAALSWGPREELTSRERIALDFAATTVDVDAVLAIRH
    ELAASGVTGAAGRLAPAAERGALLDLVLSGVTGGRVPHDQSGFRDIGNMTS
    SAVERLGRLAATGALPATLAFDVPNDALAAAYLGGRLGDTSAAGPDTPTDPAAP
    SGEPALVGMACRLPGGVTSBELWELISSGGDGI SGFTDRGWDLEHLHYHPDPHPHG
  
```

```

TSYVREGGFTHDAGDFAALFGISPREALAMPQORLLLETTWELERAGIDPFTSLRG
SRTVFAGVMYHDIISRLRTVPPELEGYSSGGSGSVSGRVSTVTFEGEPVAVSDTA
CSSLVALHLAVALRTGECTAVAGVTVMASPSYIEFSRQSGISPDGRCCKSPFAAD
ADGAAMSEGNATLLVERLSAQRGLHPVGVVGTAVNODGANGLTAPNGPSQORVI
RDALTNAQLTATDIDLEAHGTGTLGDPTEAQAVALAMYQDRLDPFWLSLAKNIGH
TQAAGVAGVMKAVLSMRHGVMPKSLHIDEPTFEDVMSAGTVELLAARDWPGADERP
RRAGSVFGVGTNAHVILEQAPAPPEPTESABEPSEAGSPGVGLBGLSPGVPV
WLSAQNETAKOLARQLVTEGETVHQIAHSLITQSTTLPHRAVLVSTREEFAVA
LEQITRTPGGKTVLFCQSGQWGMQORLWETSPLYRDTTIEIDTALSQEIEIDWSV
TDLTSQPGAPTWRDVVIOPLVFTMTALATLWQHGLTPDALDGHNSQEIIRAAATA
GAULTQDAARITTLRSKALLPLTGHGTMISVLAPEKIQELIADPDHNSQEIIRAAATA
PNTTTITGDTQATITLTLTQOLSHRIMRQPLQGVDFAGHSGLIDQLHNLQKILAPITP
QKPTIPSTYNTDNTWITDLDADYWRNLQPLVFPQAPINHLNTNNTYDITETSPHP
TLTPTQETWPNNTTITLHNNNDTHALITGALGHATGHPIITWHTLIPPTKPTLP
TYPQHTRYWLNKPTVDQATQGLGIPHPVLPVATTLAETGATILGHIPNAPHH
PWTDHANGTPLPGTALYDMALHAGHDHDTLDELITHTPTLTHPTTIQIOTISF
PDNNTHTLTHTRNSPDGPWTAHATGTLQHQHTPDPTPQTPPTWPTDAEPIPLDGTQ
RLAATGLEYGPTFGQLDMLWRDSRLFADRLPNTDHTSDAGYGIHPALLDAAHL
VAVSDGEIRLFPAMTGTILHATGATKRLVEKSWDASGSALRAFDPGQPVVTVESL
ASRVISPEQLRSASARGPLRYPEWIAWSAGATPGGAVEPERYEVPTVGDDVLADTR
RSTETLRIQOHLAADTTPLVVEATHDLAGAAVMGLIRTAQTEHPGRILIDTDNRH
PYSKALPTLIASGEPOARIEDGALITPLRTVVVAPTEPAGVEGTVLITGTSLSGLS
IARHLVTEHGRDVLVSRQGPDRAPGATELIELQQLGAQVRVSCITTDRAALALV
DEVGPFVGVHTAGALADATVDNLDALATTAAKAAHAWLHEHTDGHGLTFEMF
SSLASVLGSAQANYAANSFLDALATHRETLGLPGTSLAWSWERTSGMSAHLTRAD
HDLARAGIRPLTDEQGLEFDGALAVDLQVVTATFDLPAIGRSGAVPVLRLGLVRA
PRRAAGADSGGAEPRIETATLAPDARERSVLTSEVQAQALGHPDGPABEPDQAF
KSLGFDLTAVERFNRLSANTGLQLPATLVDPYPAAGATVLYLDDVSPAPAAAPDD
LIDEREURVLVRFVVARFVAREAGVIEALLKLAGDPDTPDAEPEDEGLDAMDTEA
LIRHVMDDTGA"
9637..14925
/note="orf2"
/codon_start=1
/transl_table=11
/product="PKS module 2"
/protein_id="AAC38062.1"
/db_xref="GI:2624949"
/translation="MTNEQLVEALRSVAKENARLRNENASLQEGAEPPVAIVGMACRL
EGVDSPESLWELMASGDVAPPTDRGMDLGLXDPDPDHLGVYVREGGFLDRAG
VFPDPPGIPSPREALAMPQORLLLETTWELERAGIDPFTLRSRGTGFTGLMYQDY
VTRDLPPAELEGYLNGSASVASGRVSTVTFEGEPAVSLDTACSSSLVALHLAVRP
ALRELLARVAGVTVMSPVHPYEFQSRGLAFDAPCKFSQQRVIRDALTDARLTAVDV
ERISDLARLGLHPVLAVRGTAVDNGASNGLTAPTGPSQQRVIRDALTDARLTAVDV
LLEAHGVTRLGDPTEAQALETATYQDQDQVWGLSKNSIKTGHITQAAAGVAGVKTVM
AMHGVMPRTLHVSEPTNVDSAGEVRLITWADWPSTQDQRPRAAVSSRGIQGTNA
HVLQEQADAPEDPVTPPAGVPVWLSAHTGALSALAEHLAAHVEEDRAPLGDVA
HTLNTATLPHRAVAVVGTREFFVAAARVVRPRRGPPLAVMTFGGSGRPGWGRQ
LYEHFPFAQALDEVFALATPGLREVMPDQAEQLQRTDHAQALAFATFALTYLME
SWGLRPMVCGHSVEITAAHVSGTLTPDAHLVTRTGLTMQLNMLNVSHAFSP
HTLQPHLDNHDITISIAINGHATVLSGDRITLHHTATLQNTKNTKMLNVSHAFSP
MQPILOPPTTLNLTTHHPHTPLISMLTATPHTDTHHTOHTAPVRYDTLHLHL
HGLITVLEIGDPTLTHARTITPLTTHLIPPTERNNEVRSNEALRGVSVCHSV
DMRALTPTRGTSPLTPYFQRDFWLDAGGAEVSGAGLGTGTDHPLGAVTVADT
GELISGLRSTSTHMTLVNVTIVPGTALLDLALHAAEHDDHTTVDESLVITPL
ALHTPTSSQSTGAETDTHRPILAHNSRDTGTWTRHTTGLSNQTHAEALSWSPT
DAHOLDITAAQOLADTHGVPAQGLHTLYQRDNTLFAELPELPAAGLPTGHLHP
ALLHAPAAVSAARDEESGILQPFVWAGTVHSTGATHRLVRDLTDQTCARRLAT
AGDILDTTHRTITAGTLRIQOHLAADTTPLVVEATHDHLHDDHLAGAVWGLIRTAQTEHP
DTTCQPVITISLTRLHPOLHAAQEPETHLITLWSPTELPVTHGHTERYVTP
AGDILDTTHRTITAGTLRIQOHLAADTTPLVVEATHDHLHDDHLAGAVWGLIRTAQTEHP
GRIITDNDNHTSRDALASVAVSGKQARIDGAILTLPRLTRUSVDFRSTRERPD
SMGVLVTGGTGGLSLIRAHVTEHGVLDLVSQGPADGAGTALTNQKQGHARI
RTISCDLTFTALTALDEIGLTVGVHTAGALADTTDHLNPDTLTTLTTPKAPAW
WLHSLTDHDLALFVPSVAVGNAGVAAANSFLDALITHRRQGLPGTSLAW
GHWREGTQAHLLQADHDMTRAGIHLTDAEGTLPDLDRLGLGAVAPVKLHLPT
LNRADTPAVLRGLVRAARQATVPAESWADRILAMPEERDRAVLDVROQASAR
VLGHSDADRIDPQGAQFKDTGFDLSATVFRNRVTTATGLRLPATLIFDHPITNALADH
LILAQGLDGGSGSPVDRLEFRERQOLFTRPDDEVGTAIVRLERSILAHLPADAT
APAARTASQVLDLTLEASQSELLSFIDKEGRASN"
14964..21266
/note="orf3"
/codon_start=1
/transl_table=11
/product="PKS module 3"

```

CDS

CDS









Streptomycineae; Streptomycetaceae; Streptomyces.

1  
Schwecke, T., Aparicio, J.F., Molnar, I., Koenig, A., Khaw, L.E.,  
Haydock, S.F., O'Leary, M., Caffrey, P., Cortes, J., Lester, J.B.,  
Boehm, G.A., Staunton, J., and Leadlay, P.F.  
The biosynthetic gene cluster for the polyketide immunosuppressant  
rapamycin  
Proc. Natl. Acad. Sci. U.S.A. 92 (17), 7839-7843 (1995)  
95372374  
7644502

2  
Molnar, I., Aparicio, J.F., Haydock, S.F., Khaw, L.E., Schwecke, T.,  
Koenig, A., Staunton, J., and Leadlay, P.F.  
Organisation of the biosynthetic gene cluster for rapamycin in  
Streptomycetes hygroscopicus: analysis of genes flanking the  
polyketide synthase  
Gene 169 (1), 1-7 (1996)  
96186895  
8635730

3  
Aparicio, J.F., Molnar, I., Schwecke, T., Koenig, A., Haydock, S.F.,  
Khaw, L.E., Staunton, J., and Leadlay, P.F.  
Organisation of the biosynthetic gene cluster for rapamycin in  
Streptomycetes hygroscopicus: analysis of the enzymatic domains in  
the modular polyketide synthase  
Gene 169 (1), 9-16 (1996)  
96186896  
8635756

4  
Haydock, S.F., Aparicio, J.F., Molnar, I., Schwecke, T., Koenig, A.,  
Marsden, A.F.A., Galloway, I.S., Staunton, J., Leadlay, P.F.,  
Boehm, G.A., Staunton, J., and Leadlay, P.F.  
Divergent sequence motifs correlated with the substrate specificity  
of (methylmalonyl)-CoA: acyl carrier protein transacylase domains  
in modular polyketide synthases  
Unpublished  
5 (bases 1 to 107379)

Direct Submission  
Submitted (06-JUN-1995) T. Schwecke, University of Cambridge, Dept  
of Biochemistry, Tennis Court Road, Cambridge CB2 1QW, UK

FEATURES  
Location/Qualifiers  
1..107379  
/organism="Streptomycetes hygroscopicus"  
/mol\_type="genomic DNA"  
/strain="NRRL 5491"  
/db\_xref="taxon:1912"  
1..107379  
/product="polyketide immunosuppressant rapamycin"  
/note="biosynthetic gene cluster"  
complement(1..871)  
/gene="orf22"  
complement(<1..871)  
/gene="orf22"  
/codon\_start=1  
/transl\_table=11  
/product="monosaccharide transporter"  
/protein\_id="CAA60449.1"  
/db\_xref="GI:987089"  
/db\_xref="GOA:Q54286"  
/db\_xref="SPTREMBL:Q54286"  
/translation="MAPRPSVRSVADVVALIDGKGTITATAPALILENPLGWLATRIGRVPPLLIKLI  
NAALSGIGPMTTOMELTSTQVLSITATAPALILENPLGWLATRIGRVPPLLIKLI  
FAIAGALLAFAGFVVLVGRVILGVAGVMDFAVALLAEITPAKIGGRNLWQAV  
WYIATNSLNLVLFNLDVGADIRWISVGSAAVAVALLGQWLKESPTWLAKG  
RLGCAITNLDKIVGKAVAGTPDEMTTPAEAPVIGLRQGRLEFRGEYLPRTILSVI  
SLGOSMQYFVAGW"  
complement(1150..2319)  
/gene="orf22"  
complement(1150..2319)  
/gene="orf22"  
/codon\_start=1  
/transl\_table=11

gene  
CDS

misc\_feature  
1..107379  
/product="polyketide immunosuppressant rapamycin"  
/note="biosynthetic gene cluster"  
complement(1..871)  
/gene="orf22"  
complement(<1..871)  
/gene="orf22"  
/codon\_start=1  
/transl\_table=11

gene  
CDS

gene  
CDS

/protein\_id="CAA60450.1"  
/db\_xref="GI:987090"  
/db\_xref="GOA:Q54287"  
/db\_xref="SPTREMBL:Q54287"  
/translation="MKELYDDSEFSFEALRAAGYAAYGADLGEVLVTCROIPEGDEE  
AWSQMAATARIERIGDDALLAGHVSAREALLASNYVXTADFYRENPAADAESA  
RLAKASQOTFADAAPLLDTPARALRIPYEDTLPGLFLVDDSGTFRPTVLFHGYS  
LTEDYLALAGALRGYVIAFDPGQGSTVREGQHFPPDWEAVVTPVDFAUTLP  
EVDABOLVMGMSLGGYLARAAPAFHRVAACVLYDGYDFHVAITAAAGRAAFTGG  
MEALMAONTMVRVVRNCRWTFGVSVDLVRAPAPYTMAGIADRTICTPTVLEEND  
OFFKQFORLDELTCQKELISFREDEGAGEHCEGAMFLPHQRTFDMLDTVLAG"  
2422..3036  
/gene="orf22"  
2422..3036  
/gene="orf22"  
/codon\_start=1  
/transl\_table=11  
/product="regulator of antibiotic transport complexes"  
/protein\_id="CAA60451.1"  
/db\_xref="GI:987091"  
/db\_xref="GOA:Q54288"  
/db\_xref="SPTREMBL:Q54288"  
/translation="MTAQKQKGRPATGGAALRQRTVEITAEAFELADAGYARMSME  
SVARAGVGKALYRRWFSKQMWTELYRGKVTDLPTPTAGALRTDLRELLTFRG  
QANPLLIARIALAGLAEASHDDALAEGLTGVTAERRAAAHALLRGADRGELPGLD  
LDLGTDLITAPLAPRVLVIQGRSDDEYLETTTNAEALRAAAR"  
/gene="orf22"  
complement(3056..3763)  
complement(3056..3763)  
/gene="orf22"  
/codon\_start=1  
/transl\_table=11  
/product="ABC-transporter"  
/protein\_id="CAA60452.1"  
/db\_xref="GI:987092"  
/db\_xref="GOA:Q54289"  
/db\_xref="SPTREMBL:Q54289"  
/translation="MTSATAPPVLRAGVSHYTSQRRRTAVLKIDIVAFERGTFYT  
ILPGSGKTLTLLSLAGLDTPTKGTITFEGDLTBLGLGRYRNOHAATFOOYNLIT  
VITALQNVTCAMEITGTPTTGSRRARALLLEKIGLDKAMATRNVLQSGQQQORVA  
IRALACVDLILPADEPTGNLDEDTAAGIIGTFRELAHEQDKCVVVVSHSQQLAAQSD  
RVTLRKGLTQGVNAR"  
complement(3770..5149)  
complement(3770..5149)  
/gene="orf22"  
/gene="orf22"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA60453.1"  
/db\_xref="GI:987093"  
/db\_xref="GOA:Q54290"  
/db\_xref="SPTREMBL:Q54290"  
/translation="MNFVKEAGLSLWSRKGRITLITATFLVISMVLAVGLINDATAQ  
AEQARRSVGAENLEADLCQLGGSELRAPRIDAATVDKVGALPOVRKYVTSYWDVRV  
LLKGGPQLADGGPKPEPGGTVMGVLDTSLLPDRFSKFTLLSGEHLTAADKGRK  
LLIIEERAKENKLGVDGDTITLGNDEKTTADFTVAGVYRDPSPATRPSYIGLSPANK  
LYGTVGLSALSSERNGLQAGKATFLQADATQAGFKDEAKKAGVAGPALDPSYGLKIDAND  
KAIROMTGPLKSIITATLAWMLAGAAVLLVNLAVKQKRGYGVLLAMGKKR  
KLIAQQALETLLVVGALAIGLSLFAPRLTQSAQALLGREATSAQHKIDSWQGPSSGS  
TGIGEGIDPNDKPEVDADPIDKITVRLDPADLATLGAIGLIGLLATAVPAGAVRLSL  
PRTILTKGK"  
complement(5184..6497)  
/gene="orf22"  
complement(5184..6497)  
/gene="orf22"  
/codon\_start=1  
/transl\_table=11  
/product="membrane transport protein"  
/protein\_id="CAA60454.1"  
/db\_xref="GI:987094"  
/db\_xref="GOA:Q54291"  
/db\_xref="SPTREMBL:Q54291"  
/translation="MNFVKRAGSLGARKSKSALLAIFVVICTLLIGGLLQAAAR

[illegible]

CDS  
complement (4572...5570)  
/gene="fkbo"  
/codon\_start=1  
/transl\_table=11  
/product="FK506 oxidase"  
/protein\_id="AAC6881.1"  
/db\_xref="GI:3798626"  
/translation="MSAPVAAQPCYCRKIGKPSDLEGETVLGVIBHGTGHTGASLVGDG  
VPHLATHTTRDEEAPAVHQAQPVPESTDDGIARWDEYLVFGVPRVPSGRVADA  
VSALYTRISGLRSGLGHPILLARTWYVSGINSANADGLEVRDFCVGRQAALDEGID  
PASMAYATGTHGGIACIATAGRGVNIENFAVTAHAYPAIGRPSRVFARAT  
WIGPDGRGLFVSATGHTHGGDTGQCEVALDNIARVWAGNRIIRRH  
LADYDLKVVYVRREDLAARRVCAARLSSTATVAVLTDTIARGDLLVEIGVVR"  
5798..28528  
/gene="fkbb"  
5798..28528  
/gene="fkbb"  
/function="FK506 polyketide backbone synthesis"  
/note="initiation and first four elongation steps"  
/codon\_start=1  
/transl\_table=11  
/product="FK506 polyketide synthase"  
/protein\_id="AAC6881.1"  
/db\_xref="GI:3798624"  
/translation="WARAESIRPLEHLLRVNAERHGRVAYTDSRAVYTGRLQURTG  
RIAGHLAAGSGVGRDRVLLGNRIETIEVYLAARAAVAYPLNPDAGDAELAHFLT  
DSGATVLTDDHLDOVRVGTDAVTLAAGRRPGCVAYEDLAGTEPPRPDDLLGLD  
EPANLTYSGTTPRGVVSQAORSGLWATYCDVPSMELTEDDELLMPLAPLSLGH  
LCILLAVTGASARILGCVARDVLDALAEHCTVLVGVPTWRYLLGAVSEPQTRA  
LRLVAGSTSPASLTAFEAAGFVPLLTGTGTTGAQTANTLTDTVPVSGCLFV  
PGLSRFVDPVADVRDEGELWASGPSMLGYHAQTEATAQVLADQVGTGDIAR  
QAEHTVITVGTGKELIRGENIHPREIESVAQVETVPTASGRVRLGELIPVL  
VYVPEGVVPAADLACBRQLAYFKVPDEILLHTVPTVPSGRVRLGELIPVL  
GTSGEALCELWERRLEPDTVSQVAVTRAVGIGPAEVPDQALRDEALVRQ  
AEPGSLVLTDTTDTNANGDEADELVAAALGCEQVQALRSQAVTVLVRADT  
TLPASQWALRPASPTKDLVAPSDAPPRPLAAGEVRVRAAGNFRDVLILAUT  
YFGDGAAGEAGTVEVGEVDVLPVGAQVFLVQDAFRSVAARLRLVARI PGWS  
FVPAASVTVFAATVLDAGQVRPGQKVLHAATGCVGSAAQRLARHMGADVATA  
SPAKHLLADGDAOHVANSRTTAFADAPFPVDVVLNSLTGEFLDASVELLAPGRF  
IEMGTDJPHDQPPDLSDVPDARLEULELLELFDRLDELPLVRPMDIARADA  
FGMSQARTNKTLVTLPLPGDPAVLVTGEGADAVAGLRLDETGRTHVSSAAARE  
RAIEGEGVNSATLHVLIGDPAGADAYARSASEGLPIVTVHGLPGIEQAALTGELI  
EAEIEGGSVYTVRSGAGARAPAMAGATPPLISGLAGPTGPAEPGAAGPEWGSRLAA  
ADARENLLDLVRDSVATVGLPGAEACAPRTFENGILSLITVEFAFNTVAARTGL  
RYPASTAFOHPTPRSPAHLAGSVAAPAAKALGSGEPVALVGMACRLPGVTSRGL  
WRLVASGTDAITEFPDRCWDVDELVDPPDAVGRSTRHGGFLTEATFQDAARFGIS  
PNEALAMPQOORLVLTSEWAEAFHAGIVPTLRESITGVPMGAFHGYGACRGLGIG  
VTATQSVLSGRUSYFTLGQGPVTVITACSSSLVALHQAAQLRSRGCSSALAGGV  
VMTPGSFVFFSRGLSPDRCCKFADADGTGPAEGVGLVVERQSDAERNHTVL  
AVVPGSAVQDQASGLSPNGVAQQRVIRQALANAGLSGADVDAVEAHRTGTVLGP  
IEAQVALSYGQREVPPLIGSVKSNVGHVTAAGVAGVTKMVMAMRHGLLPKTLHVD  
EPSSHVDWAGAEVVTETPWPESGARRAGVSSFGVSGTNAHVTLGVAESVRDG  
GSSGLVLPVSARTESLALOVERLGEYVRGGADAAVADGLVRGAVFDRRAVLIGE  
SAGVAVAGERTVFPVPGGSGVGMGRELMGASEVFAARMRECAVLEPHPTGWDLL  
DVLGEVAVRDEVTVPAGSAVAVSLAALQAHVPPVDAVVGHSQEEIAACVAGALS  
LEDAARVALRSQALIRLAGRAMASIAVPASAVETVGVGMIAAENGESTVWAGDP  
AAVERLYARVAGVRVIRIADYASHTHVEAIEAQLADALEGITSSTPSVFWGTV  
DSGWTEPBGDAVYVENLQOPVAMDTAVSELDSGLFIECSAHPVLLPALDQERTVASL  
RTDDGWDFRLLAALQANTQGVDDNTLIEPAHRLVDLTPTTPPHKRYWLOPAPVT  
GAGIHGFUSSAALPGSDGVLITGKLSLATHFWLNDVTQVEVLLPFGTAFIELVVRA  
GVEGCDTIDELVETPSHPASGAVDLTVTVDPDDTGRHPVSHVAREPDTWTWRH  
ATGLGRTEDTAPFNSFPWPPTGAQPSLDEFERLAGAGTVYGFAGFQGLAAWRA  
GDTVYARVADLGDADRFGVHPALLDAAHLHAGRLDAGDELELFPFSWTVGRVHAATGAA  
AVYVHRSRGAAGVAVQVADLDGRPVSVDAVLVLRPSVAPSGFNILGLEWLVHATYH  
DGADELPEGTLITATPDDDDDTPNPHNTPTRTHTQTRVLTALQHLHTNTNTHY  
VHTTDPGGAAGTGLTRTAQNEHPRHLEIETHPHPTPLTLQTLATHPHPLRTNNT  
LHTPHLPTITHTTTTTTPTNPNHAILHTTGGSGTLAGLHSLHPLTYLISR  
TTPPTTPRTHPCDLTDPQITQALTYHPOPLTFTHTAAHDDADLTNLTPOHLHT  
TLQPKADAHLHHTONQPLTHEVLYSSAAATLGSFGQANYAANAFLDALATHHT  
QGPATTIANGMHTTTTSLTQDTERQQVRDGRPLTHEAKGNHFLDASLATDAFPV  
VAAIPAPARRRVERRTARGDSGRDLAVVCVATAVLAHLHADASEIGAATAFKDLG  
IDSLSGIRLNRLENSAATVAFDHPPTDPAALARLHAGLBEIRDSDSQGAAAAA  
PGTVPTVAAAGDEPLAIVAMACRMPGCVDTPEDLWSLVESGGGDAITEFFPTDRWDLAA

gene  
CDS

Query Match 36.8%; Score 184.2; DB 1; Length 28732;  
Best Local Similarity 61.9%; Pred. No. 3.4e-13;  
Matches 310; Conservative 0; Mismatches 188; Indels 3; Gaps 1;  
Qy 1 ATGTTTCAAGTCGCGGCGACGACCGGCGGCCCAAGCGGCTGTGTCTCCACGACGCCAAC 60  
Db 6290 ATGCTCTACAGCTCCGGAACCAACCGGCGCGCCGAGCGGCGTGTCTCCGACACGCGCAGT 6349  
Qy 61 TGCCTCTGTCCTCGCTCTCTGCTACGTCGCTCTCCCGGGTGTTCGACACGACGACCGG 120  
Db 6350 GGCCTGTGTCGCGACGACTGCGACGTCGCGCTCTCTGGAGCTTACCGAGGACGACGAA 6409  
Qy 121 GTGCTCTGCGCGCTCCCGCTGTTTCCACAGCCTTTCGACATCGCCTCGCTCTGTCGCGC 180

Db 6410 CTGCTCTGGCCCGCCGCTGTTCCACAGCTCGGACCATCTCTGCTGCTCGGCTC 6469  
QY 181 ACCGTGGTGGGCGACAGCTCGGATCGCGACCGGACGCTCCCGCGACGATGATCGG 240  
Db 6470 CTCACGCTCGGCGCTCGGCGCGATCTCTCGGCGGCTTCGTCGCGCGTGCCTCGAC 6529  
QY 241 CTGATCGAGCGGAGAGCTCGACTTCTCTGCGCGGCTCGCGACCTACACCACTG 300  
Db 6530 GCCCTGGCCGAACACCCCTGCACCGCTGCTGCGCGTGCAGATGACCGTACCTC 6589  
QY 301 GTGCGGCGCCCGCGGACGCGGCTTCTCGCGCGCGAGCTCGGCGATCGGCGGCGG 360  
Db 6590 CTCGCTGCGG---TGTCGGGGAACCGCAGACCGCGCGCTCGCGGCTGGTGGCG 6646  
QY 361 GCGCGGCTCTCGGCGCGCGGCTGCGAGGAGTTCGAAAGACCTTCGCGGCTCCCGCTG 420  
Db 6647 GGGTCCACCTCACCGGCTCGCTCACGCGGCGCTTCGAAAGCGGCTTCGCGGCTGCGGCTG 6706  
QY 421 ATGACCGCTACCGCAGCAGCGGACCTGCGGCGGCTGATCACCATGACCGCGGACGCG 480  
Db 6707 CTGACACCTACGCTGCGACCGAGACCAACCGGATCGCTCACCGGCAACACCCCTGACCGAC 6766  
QY 481 GCCCGCTCGAGGCTCGTGC 501  
Db 6767 ACCAGGCTCCCGGCTCGTGC 6787

RESULT 11  
AF235504/c 77534 bp DNA linear BCT 19-JUL-2000  
LOCUS Streptomyces hygroscopicus var. ascomyceticus FK520 biosynthetic  
DEFINITION gene cluster, partial sequence.  
ACCESSION AF235504  
VERSION AF235504.1 GI:9280381  
KEYWORDS Streptomyces hygroscopicus subsp. ascomyceticus  
SOURCE Streptomyces hygroscopicus subsp. ascomyceticus  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomycetes.  
REFERENCE 1 (bases 1 to 77534)  
Wu, K., Chung, L., Revill, W.P., Katz, L. and Reeves, C.D.  
AUTHORS The FK520 gene cluster of Streptomyces hygroscopicus var.  
TITLE ascomyceticus (ATCC 14891) contains genes for biosynthesis of  
unusual polyketide extender units  
JOURNAL Gene 251 (1), 81-90 (2000)  
MEDLINE 10863099  
PUBMED  
REFERENCE 2 (bases 1 to 77534)  
Reeves, C.D., Wu, K., Chung, L., Revill, P. and Katz, L.  
AUTHORS Direct Submission  
TITLE Submitted (16-FEB-2000) Kusan Biosciences Inc, 3832 Bay Center  
JOURNAL Place, Hayward, CA 94545, USA  
FEATURES Location/Qualifiers  
1. 77534  
/organism="Streptomyces hygroscopicus subsp.  
ascomyceticus"  
/mol\_type="genomic DNA"  
/variety="ascomyceticus"  
/db\_xref="taxon:132248"  
complement(412..1836)  
/gene="fkbw"  
complement(412..1836)  
/gene="fkbw"  
/codon\_start=1  
/transl\_table=11  
/product="pkbW"  
/protein\_id="AAF86380.1"  
/db\_xref="GI:9280382"  
/translation="WVLPIGVGRAPRRRTGLALTLAAVALLPSPATARPAPDAPDG  
HCARLSQRVFAAARQQADCLDELTTAGTAATGCTDPAWAGLTGPKNLATPTGVPGIQ  
IDGYPFDTSTTNHGMNDAOFVIRLPDRMNGGLVAGTGCNREQANDRAIAADWL  
ARGYAAATDKNTGTPAFVRDGRPGDAVAEWNTLITQLTRAARASVAQRVHRPLERT

source  
gene  
CDS

gene  
CDS  
complement(2020..3579)  
/gene="fkbU"  
complement(2020..3579)  
/gene="fkbU"  
/codon\_start=1  
/transl\_table=11  
/product="polyhydroxybutyrate depolymerase"  
/protein\_id="AAF86381.1"  
/db\_xref="GI:9280383"  
/translation="MOPPPRGLITPLPLSSPPVGSLSRPGRRGVLTRIVAVVALV  
LGAALGPATAHAAGLAKGLTKADLTEVADFGTNPRLNMVYVVPASLPAPAVV  
FAUHGCTQDQAGADNSGLSPADRYGFLVFAETSSNNANCFNWFQSSDNERGQG  
EASIRQMAHTVAYGADPQRTYITGLSAGGAMTSVMLATYPDVFOGAVVAGLPFG  
CATDVSAYLCMNPFGDITADQWARRVDGYPWSGWPWRAIWHGDKDQTVAPRAD  
ELRDQWAVHGVSQTPDRTSIVGPNSTRHEYLADSGVAVENRVPGI GHGTPVDFG  
TGAOCGSGTAAYFLDSICSSYWTOTFGLSGSADPGSLPAPSLAATGATDTTISL  
TWKPDGATYAVHRDQAOITTSATSYTDTGLRAGTSHTYAARADADGAPLSGA  
VTAGTTGATVACWTAAGYAHVQAGRAITSAGTYIYAKSGQMGSLYNTFTVTLKBSPT  
GYFTVANDTCP"  
complement(3969..4496)  
/gene="fkbR2"  
complement(3969..4496)  
/gene="fkbR2"  
/note="transcription regulator"  
/codon\_start=1  
/transl\_table=11  
/product="Fkbr2"  
/protein\_id="AAF86382.1"  
/db\_xref="GI:9280384"  
/translation="WVGQGWGRYFEDFTGQVYRHPIGRTVLETDNSWLTLLTONTAP  
LHFDHYSGAGTGWKPLVDSTFTLALVTGQSVTDISQHWANLQWDRVRLNPNFEGD  
TIYSQSEVLNTRSAASRPDVGIIISVRTIGYNQDGIIVITFERTILMVYRGGHPRFAAV  
EPWDESRRAAGLH"  
complement(4595..5488)  
/gene="fkbR1"  
complement(4595..5488)  
/gene="fkbR1"  
/note="transcription regulator"  
/codon\_start=1  
/transl\_table=11  
/product="Fkbr1"  
/protein\_id="AAF86383.1"  
/db\_xref="GI:9280385"  
/translation="MEIRELQWFTALAEHEHVTITAERLNI SQPTLSAIRRLRRVVG  
VPLFDRHONRLNLYKGFVFRHAHRAISEVGAQRITALIDPDTGTVALGELHYSY  
TMDVLLAGYRRLAPGTTPELRGDAADAVDDVNSRLDGLTSPRPAAGDVEHTPL  
RDEALCLVPPGHRLLAPRRVRVTAELADEAFVALEPVGRLQITDRLCAAAGFVPRV  
LESTELSTLVAALGAVAVAPLITGSPSGPDHNAVPLDDPQARRVGVVTLAGGP  
RAPVVARFYDVRSRAHAPHV"  
5601..6818  
/gene="fkbR"  
5601..6818  
/gene="fkbR"  
/note="L-carnitine dehydratase"  
/codon\_start=1  
/transl\_table=11  
/product="Fkbz"  
/protein\_id="AAF86384.1"  
/db\_xref="GI:9280386"  
/translation="MSEPHRPPQERPRAGPLISGLLVLSLEQAVAPAPFATRIHLADIGAR  
VIKTERPGSLARGYDRTVRGMSHFVNLARGKESVQDVRSPGNRHRLHRLVQDRA  
VLVQNLAPGAARLASATRSRGATEADHLRTYATAPVAPATADKADYDLVQCEAGL  
VSTICTPTPSKVLGIADI CAGWYSGILITALLIKRATGRGSLQLEWFMLEALGEMW  
GYAEYTVGTGAPARACASHTAPYGFPTTRDQGTINLGLONREWASFCGVVQOR  
PGLCDDPPSGNADRVARTELDLYVSEVTGTLGEEELVAREASIAVARQTVREF  
SEHPLODRGWAPDPSVPGALEGLIPPTVTHGHEPRRLGRVPELGEHTESVLAMLA  
PHSADREAGHAE"  
6808..8052  
gene

```
/gene="fkbF"
/locus="8052"
/translation="similar to Rhizobium trifolii MatC; dicarboxylate carrier"
/codon_start=1
/transl_table=11
/product="fkbF"
/protein_id="AAF86385.1"
/db_xref="GI:9280387"
/translation="MENEUTGVLLIAAVFLLAGVRLNGLLALVATFLLGVVLDRT
PDEVLGAPASFLVAVTFLEGIARVNGTVDWLVRAVAVGAVGVAVPWLFLGLA
ALLCATGAAPAAVAIVAPISVAFVHRHIDPLVAGLMAVNGAAGSAPSGILGIV
HSALEKNHLPVSGGLVFACTFAPNLAFAVAVSMLVLRRLRLEPHDLDDTDPEGPAS
RGAEHMTLTAAALVLTGTVLSLDTGFLATLAAALLPFTSQOATKIAFPV
LLVCGIVTVALLQELGIVDSLGKMTAAGTLPILLAAVLVCYGVGVSAFSTGILGA
LMLPEPFLKSGAIGTGMVMAAAATVVDASPFSTNGALVAVANAPRLRPVYQGL
LWAGVGCALAPAAVAAVFA"
/gene="fkbG"
/locus="8824"
/translation="O-methyl transferase"
/codon_start=1
/transl_table=11
/product="fkbG"
/protein_id="AAF86386.1"
/db_xref="GI:9280388"
/translation="MANITLSDTLAYVRKVSRLRDEVLVSLRQAETAEPLGGVULPV
QAEQGFLEFLVRLDVGARTVLTGISTICLARGAPGGRVTVCDVMPKPEVGE
RYWEGAVAGVLDVIRGARTVLTGLLDEAGAPSPDFMVFIDAKGAPPAYEALP
LVRRGLIVDNTLFRGVADEAVQPDVAVRELNAALRDDDDRLAMLTADGVTL
LKRK"
complement(8802..8884)
/gene="fkbH"
/locus="8884"
/translation="O-methyl transferase"
/codon_start=1
/transl_table=11
/product="fkbH"
/protein_id="AAF86387.1"
/db_xref="GI:9280389"
/translation="MTIVKLVMDLNTLMRGTVLEDDDEVLTDEIREVITLDDRGI
LOAVSKNDHDLAWERLGVAEYFVLARIGMPSQSVREIATELNAPFTTIAFID
DOAEAEVAFHLPVRCYPAEOATLLSLPESPVPSTVDSRRRLMYOAGFAPDQA
RAYSGPDEDFLSLDLSTIAPAGEEELSRVZELTLRTSQMNATGVHYSDADLRACF
TDPAEVLVVTGDRFGPHGAVGIIILEKPSFTWHLKLTATCRVVSFGAGATILNWL
TDQARAGALHVADEFRTDRNRNMEIAYRFAGFADSDPCVSEVAGASAAGVERLHLE
PSRAPPTTITLTAADIAFVTVSAG"
complement(9895..10995)
/gene="fkbI"
/locus="10995"
/translation="acyl CoA dehydrogenase"
/codon_start=1
/transl_table=11
/product="fkbI"
/protein_id="AAF86388.1"
/db_xref="GI:9280390"
/translation="MFERDALLTDLVGDRAAEWNTSGELPRDLLVRLGADGLLCAEVA
AEHGLLGSRENGEFTAHVGLSCLSLRSYMTSQGMAWTVQRLGDAGQRTFLKEIT
SGKLAAGFGEROAGSLMSTRVLDGTVAVDGHKVTWTAAYADHLVVFGLQED
GSAVAVVPADTPQVRYVPKPSGCAAGADHLIDQVVPVAGVLAGASLPMIV
AASLARYKSVANGCGVILRACRTAAVAHARTREQGRPLGDHLVAGHIALDTAEQ
IAARVCBSADHWDGSPWPATILAKHVAERAAAGAAATAAQVLIASAGAREGHVVE
RAYDAKIMEIIEGSEMCVMLAQHALALPA"
complement(10988..11248)
/gene="fkbJ"
/locus="11248"
/translation="acyl CoA dehydrogenase"
/codon_start=1
/transl_table=11
```

```
/product="fkbJ"
/protein_id="AAF86389.1"
/db_xref="GI:9280391"
/translation="MTAPVKDDPVAHALVAELKTKTRSDMPVDRLFAEGGLTSLFAM
ELVVLEKTFDVTIAGPDLQANFRTVESMVALVRLRAVDA"
complement(11245..12093)
/gene="fkbK"
/locus="12093"
/translation="acyl CoA dehydrogenase"
/codon_start=1
```

Query Match	36.8%; Score 184.2; DB 1; Length 77534;
Best Local Similarity	61.9%; Pred. No. 2.5e-13;
Matches 310;	Conservative 0; Mismatches 188; Indels 3; Gaps 1;
QY 1	ATGTTCTACACGTGGGACGACGCGGCGCCCAAGGGCGTGTGTCCACGACGCGCAAC 60
Db 46079	ATGCTCTACACCTCCGGCACCAACCGCGCTCCCAAGGGCGTGTGTCCACGACGCGC 46020
QY 61	TGCTCTGTGTCGTGCTTCTTCTGCTACGTGCGGTTCCCGGGTGTTCGGACGACGACCGG 120
Db 46019	GGCTGTGTGTCGGATGACATGCGACGTGCGCTCTCTGGCGGTGACCGAGGACGACGAA 45960
QY 121	GTGCTCTGGCCGCTCCGCTGTTCACAGCCTTTCCACATCGCTGCTGCTGCTGCTGCTG 180
Db 45959	CTGCTCTGGCCGCTCCGCTGTTCACAGCCTTTCCACATCGCTGCTGCTGCTGCTGCTG 45900
QY 181	ACGTGTGTGGGCGCAGCTCGGATCGCGAGCGGCGAGTCCGCGACGACGACGACGACG 240
Db 45899	CTACGCTGTGGCGCGTCCGCGCTTCTCTGGCGGCTTCTGCTGCGGCTGCTGCTGCTG 45840
QY 241	CTGATCGAGCGGAGAGCTCGACCTTCTCTGGCGGCTGCGGACCTTACACCACTACCA 300
Db 45839	GCCTGGCGCAACATCTCCAGACCGGTCTGCTGGCGGTGCGGACGATGATACCGCTAC 45782
QY 301	GTGCGGCGCGCGCGCAGCGCGGTTTCTCCGCGCGAGCTGCGGATTCGCGCTTGGCGGG 360
Db 45781	-TCCTCGCGCGCGGTGCTCGCGGCGCGGCTGCGCGCTGCGCGCTGCTGCTGCTGCTG 45723
QY 361	GGCGGCGTCTTGGCGCGCGGCTGCGAAGCGAGTTCGAGAGACCTTCCGGGTCCCGCTG 420
Db 45722	GGATCCACCTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 45663
QY 421	ATCGAGCGCTACGCGGACCGAGACCTGCGGGGCGATCACCATGAACCGCGGACGCGC 480
Db 45662	CTGACACGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 45603
QY 481	GCCCGCTGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 501
Db 45602	GCGCGGCTTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 45582

RESULT 12  
BD235937/c  
LOCUS  
DEFINITION Polyketide synthase and recombinant DNA construction thereof.  
ACCESSION BD235937  
VERSION BD235937.1 GI:33045707  
KEYWORDS JP 2002526114-A/1.  
SOURCE Streptomyces hygroscopicus  
ORGANISM Streptomyces hygroscopicus  
REFERENCE 1 (bases 1 to 77536)  
AUTHORS Reeves, C., Chu, D., Khosla, C., Santi, D. and Wu, K.  
TITLE Polyketide synthase and recombinant DNA construction thereof  
JOURNAL Patent: JP 2002526114-A 1 20-AUG-2002;  
KOSAN BIOSCIENCES INC  
COMMENT OS Streptomyces hygroscopicus  
PN JP 2002526114-A/1  
PD 20-AUG-2002  
PF 01-OCT-1999 JP 2000574696

```

FEATURES
  source
    Location/Qualifiers
      1..77536
        /organism="unknown"
        /mol_type="genomic DNA"

ORIGIN
  Query Match      36.8%;   Score 184.2;   DB 6;   Length 77536;
  Best Local Similarity 61.9%;   Pred. No. 2.5e-13;
  Matches 310: Conservative 0; Mismatches 18;
  Indels 3;   Gaps 1;

```

Accession	Sequence	Length
QY	1 ATGTTCTACGTTCGGGACGACCGGGGGCCCAAGGCGTCGTCTCCACGACGGCAAC	60
DB	46081 ATGCTCTACACCTTCGGGACCAACCGGCGTCCCAAGGCGTGGTTTCGACACGGCGAC	46022
OY	61 TGCCCTCTGTCGTTCGTTCTCTGCTACGTGCGCTTCCCGGGGTCTCGGACACGACCGG	120

QY 121 GTGCTGTGGCCGCCTTTCACAGCCTTTCCGCAATCGCCTCGCTCTCTCCGCC 180

181	ACCCTGGTTCGGGGCAGCGTTCGGATCCCGACGGCAGCTCCGCGCAGCAGCTGATGCGG	240
45901	CTCAGCTCGGCGCTCGCCCGTATCTGGCGGCTTCGTCGGCGCGTACGCTCTCGAC	45842
241	CTGATCGAGGCGGAGAGCTCGACCTTCTTGGCGGCGTGGCGCAGCACCTTACCAACCACTG	300
45841	GCCCTGGCGCGAACACTCCAGCACCGTCTCGTCGGCGTGGCGACGATGTACCGTACC--	45784
301	GTGCGGCGCGCCCGGAGCGCGGTTCCTCGGCGCGAGCGCTGCGGATCGGCTTCGCGCGG	360

	45783	45784	45785	45786	45787	45788	45789	45790	45791	45792
Db	-TCC	TGGG	CGCG	TGTCCG	CGAC	CGCG	CGCTCG	CGCG	CGCTCG	CGCG
QY	GGCG	GGGT	CCTCGG	CGCG	CGGGCT	CGGA	AGGAT	TCCAG	GAGAC	CTTCGGGGT

Accession	Gene	Accession	Gene
45724	GGATCCACCTCGCCGGCGTCGCTCACCAAGGATTTTCGAGCGCATGTTCCGCGTGCCTCG	45665	
Db		Ov	
421	ATCGACGCTTACGGCAGTACCGGAGACTTCGCGGGCGATCACCATGAAACCGCCGGACGGC	480	

45664 C T C G A C A C G T A C G G T G C A C C G A G A C G A C G G G T C G C T C A C C G C C A C A C C C T G A G G A T 45605

QY 481 GCCCGGCTCAGGGCTCGTGC 501  
45604 GCGCGGTTCCCGGCTCGTGC 45584

RESULT 14  
AF210249/C

LOCUS	AF210249	77457 bp	DNA linear	BCT 29-AUG-2000
DEFINITION	Streptomyces verticillus bleomycin biosynthetic gene cluster, complete cds.			

ACCESSION  
AF210249 AFI49091  
VERSION  
AF210249.1 GI:9937210  
KEYWORDS  
SOURCE  
Streptomyces verticillus

ORGANISM  
*Sreptomyces verticillus*  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE  
1 (bases 74421 to 74693)  
AUTHORS  
Du, L. and Shen, B.  
TITLE  
Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer *Streptomyces verticillius* ATCC 27861

JOURNAL  
Chem. Biol. 6 (8), 507-517 (1999)  
MEDLINE  
99352421

10421758 PUBMED  
2 (bases 1 to 77457)  
REFERENCE  
AUTHORS Du, L.; Chen, M.; Sanchez, C. and Shen, B.

**TITLE** An oxidation domain in the BLM11 non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in *Streptomyces verticillus* ATCC15003

**JOURNAL** FEMS Microbiol. Lett. 189 (2): 171-175 (2000)

**Table 1**



MEDLINE  
PUBMED  
10930733  
3 (bases 1 to 77457)  
Du,L., Sanchez,C., Chen,M., Edwards,D.J. and Shen,B.  
The biosynthetic gene cluster for the antitumor drug bleomycin from  
Streptomyces verticillus ATCC15003 supporting functional  
interactions between nonribosomal peptide synthetases and a  
polyketide synthase  
Chen, Biol. 7 (8), 623-642 (2000)  
20500448  
MEDLINE  
PUBMED  
11048953  
4 (bases 1 to 77457)  
Du,L., Sanchez,C., Chen,M., Edwards,D.J. and Shen,B.  
Direct Submission  
Submitted (01-DEC-1999) Chemistry Department, University of  
California at Davis, One Shields Avenue, Davis, CA 95616, USA  
On Aug 29, 2000 this sequence version replaced gi:5328869.  
Location/Qualifiers  
1. .77457  
/organism="Streptomyces verticillus"  
/mol\_type="genomic DNA"  
/strain="ATCC15003"  
/db\_xref="ATCC:15003"  
/db\_xref="taxon:29309"  
complement (1199. .2185)  
/codon start=1  
/transl\_table=11  
/product="Syrp-like protein"  
/protein\_id="AG02342.1"  
/db\_xref="GI:9937211"  
/translation="MTGSVTLTPLGGIIPRPRGEGTGAEDYDLGDLGADGPDVWRHA  
GTLERLATDGLILHGLDPTDGDVGDHVVGSGLDPLPYTERSTPRSVVKGNIY  
TSTEPADQPIPMENNSVAHWPSTLYFFCHAPTGTGATPIADRAVLIDLPAEVR  
RESQGVYTRFADMGISQEAQFQEDRGVVERCHRAHQEFSDGDLVLRHRP  
ATVPGTGAEFVFNQAHLEFSSLDPLDRQVILETYGNGLPRDLFADGTPIDAD  
LATVRAATRAALDPREGDMLVNDLNRMAHREFFTGERRVLVAMTSADS"  
complement (2182. .3918)  
/codon start=1  
/transl\_table=11  
/product="peptide synthetase NRPS12"  
/protein\_id="AG02343.1"  
/db\_xref="GI:9937212"  
/translation="MTLIRDAFLDQARTPDAAHVHVGDTVMVRELELRAGRMART  
LILAEHPSRDLTLPDEALAPARPDAAVVRAGDPAYVYITSGSSRPGKVLVEQGA  
LAAGAGPGLTAVARLPGRGPEVAALLAVLITAGVYPLADDDPPDRCSHILDDCAA  
LHAGPQARVDGLSGRTVLHSSLSFDMAVTSLMGLVSGGAHVLDLKALASGTOP  
PAAARPSFLKVPFHLGLDPLDSCLPDGLVIGGALTGSAALGPRWAHPDVTY  
VNEYGTEATVPGCCATVPGDAVDPGAVPIGRPFAGTRLYLDADGEPVAVGGVGL  
HIAGDQARGYLGRPLTEERFVDPDFAADGSMYRTGDLVRRPDPDGLYLGADGQ  
VKVSGYRIEPEETEAVLRGHAGVRDCAVAVGEADARLVAIVVYVDPDPPGTAAPAR  
HAAALPPYVVPATVTVPELPLTPNGKLDRLDAPGPAGDAGPGRTPAETLLCELL  
ARALGPEIDADADFLTSGTSTITALKLAVAGRRVGIRLELITVLEARTVRIILAAQP  
DAASPLAEVPE"  
complement (3915. .4733)  
/codon start=1  
/evidence="not experimental"  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AG02344.1"  
/db\_xref="GI:9937213"  
/translation="MSKNAAHWSRIKGTADPGVVLAVDFYGTGQREATFPHLDDLTL  
PVMVWAPVAPDGDSTAGHAKRWITGDLTVLAGRVRAIVGVCAGVFPASALA  
DALVREGRHPRVLYNPSAPGATVLRDPRGLIAGMDLLTDGERAALIAETVAIRRA  
HAPDLPVPAERVAALYREGDILLCELGVDSAGFAELAVLHSLAYLTAALDVPT  
FLWRGAVSLTRESHQGDFDVEHGFDAARAILSSPQVVAULTALLREHAR"  
complement (4741. .6024)  
/codon start=1  
/transl\_table=11  
/product="putative transporter"  
/protein\_id="AG02345.1"  
/db\_xref="GI:9937214"  
/translation="MTARVDAATPTTYAVLAVREARAPLISGSLARMSFAVLPALILIS  
VRDATGSFAVAGLTSGALSATLTFAPARARLIDRRGSRGLVRLTVPYLLGLAVLIT

LAEEAPTALLVAAAVAGVFAPPLGPTMRVLMARILHGROPLHHTAALDSVTVEVV  
FTVGLLAGLIAVAAPLASMITVMVLTIAAGTACFVLSAATAAPASGEADEDRPHGR  
PMALPGMRTIVLSFGGVGLVGVVLQVLPFTADHAGSPGAGGILLISLISAGSAGVGLA  
XGRJAMSTPVVRVLTGFTLAVPLCLTASVPVPAFAFALLVGLCLAPLFTTAYLL  
VNDIVTASGAPTANTWSTANNGGFAAGSAAAGVLLDSRGPTVTVTAAFAVAATA  
VMTVIRRTLLIGAGHPPEAAATPADRTAPAEAE"  
complement (6021. .6773)  
/codon start=1  
/transl\_table=11  
/product="thioesterase"  
/protein\_id="AG02346.1"  
/db\_xref="GI:9937215"  
/translation="WDFELTRVNMFWFSGCDGRPRVRLCALPYAGGTAAVFKDWPAAL  
PFGVELLTAHLFGGRDFTPEPPATLEATEALCEALPFDLPTVVLHSGMGLGIE  
VAARLAARGRAPNLLIAACRPPHPVDPDASGPTVEAEALATLRAERPDTRDELDM  
EAVLPALVADITAGDRVPRPRPLDLKVVIGADDDGTDWRTTLGWRACTARDCEV  
VVLPGHYFLETDRRAVLTTRYATDLAEVGA"  
complement (6784. .7797)  
/codon start=1  
/transl\_table=11  
/product="clavaminate acid synthase-like protein"  
/protein\_id="AG02347.1"  
/db\_xref="GI:9937216"  
/translation="WIDJPLRTVALTGESAEVDDLLRLTLADVPYDSTVGLHRELA  
AQELPLIRAEITGMRLYDSPRALVVTGVDDEIRIGPTPAARPADPDERKDLLELL  
LIHAALGEAFGWATQONGRLVHVLVPVGEETAQMGSSSETELLWHTEADHPLRCL  
YVGLICLRNHORAAATVGVWPLDRLTTRDRAVLEPRYLIRPTDTHAQNATGRSA  
ERFAIAEMDDAPERVAVLFGDPEPYLIRPAYMSPAPGAAARAVDVTALLIDE  
LRHVLDAGSLLLVDNYQAVHGKPFAPAAAYDGRDLWLKVNITRDLRSRSARRSATS  
LIV"  
complement (7859. .9214)  
/codon start=1  
/transl\_table=11  
/product="putative transporter"  
/protein\_id="AG02348.1"  
/db\_xref="GI:9937217"  
/translation="MWSRLRDLRQLGGEDPVVAQESPDQAGTTCQDGDGLNVAAL  
AALREVEVPRPLPHHAGIRHVEHSSAAVALGPMQVASTTGVAGDGHQAGNRGQ  
VTGVGVREAVGDELGAEDPHARQRLDLDLMMARLADLVNLTQVVGQDL  
RGQVSHDLGDDVLPQGLSLGSLQRRRGDVGTAHASVQGPQGPQGPAAAEASCRG  
LYARQQDORALARAAYVEGFORGEDAGORVATDHPDPVGHVGVGQVQGVQVQGL  
GHRDRREVRVAGGDDVGVAGVGLRLAPVSAHVADRVAGVGLLAVGVEQSQ  
QGGAGRVHVCVPLVGVQVEDLTHGGDRLLVADLLRFQRTGLVDHRRPVNPLARY  
H"  
complement (9211. .17736)  
/codon start=1  
/transl\_table=11  
/product="peptide synthetase NRPS11-10"  
/protein\_id="AG02349.1"  
/db\_xref="GI:9937218"  
/translation="MASDALSAHLEVSDCRQGGITVINDGSELQAQVAAVAFERY  
AIAPBETA LHKGAATGYDELNREBELTATRLADAGAPSTLVAVALPDPDLVATLC  
ALLKGAACPLDIPGIPACRLREIMADAPDVLVYTRAVAPATGGVPLFLDADPT  
CSAVLPBSAGTASEIAYVLTPTTPEKSENSVSTRDMARLDDPTAGIPARAELIR  
LVAPLSGGRLVDADETPRVTPREAPDMVEDVVAQVCAVLGVDRVGRDRKFFDL  
VGRSEPLPSFAQORLFLDRMPDRAFTVCMDFRVCIDIGALRRALRLVAGRHE  
TIRTAVERDGVYQLVGPADGGAFAVATPVDLSLELPEAREEAVNLVAAEART  
PFRADGALLRVVARLADDDHVLVSTHIIYSDANSVGLVDELGLYECVYGDPA  
ALPPPQVADFAVQWAWAGVQEEHLYWKRALDAGAPSVLRLPMDHPRPAVQSER  
GETVGPALPDVAALKEKREGQATLFMTLIGAFQVLLARHAGQDIIVGVFAAOKT  
RDETPLGVFVNTPLRAICAPGLSFDLDDQVREALGAFADQLPFEALVEALAP  
EDLGHNLVQVTFOLLGTPAARPDILIGEVERYVQEAQVSDFLSLDKRADDSYR  
GIINYCPDLFDRRMEVLYGHVLTILGAAADGPPIGELPLSDGAEARLILGAGVRE  
DRAVAGSVPRPFAEVATDARAVTCATLTFAELNDRVLRQAQLIGAGVIRE  
TPVAVLRPSTDSVVALAVMRAGGVIVELDDPMDADRTAYLILDDTAASVITRDLPA  
LFGRLHVDPRPAADGLVPAPRIDPQAAYVITSGTAPKGVVVRHRSNLHLSAL  
QATFLGHPYLAGADGVPDCAKLRITTLTAPFTFDASMEQLSWMGLHGFIFVDEVR  
RQPSALVRVVRREHIDVITTSQELSLVSHGLLDGEWAPSMVMVGVEAVSPSLMRTL  
RQORFRCNLYGPTAETVDCDLSDDADVPVIGTLPHTHVRVLDRLRPVGV  
AGIYUGGGLGALYLNRLNALTARRVADPYPTGSRLLYRTGDRARWRPDGTLVYG  
RTDDQIKINGFRVPEGEIEAVLTHHPAEEAVVDDAHARLVAIVTLAEGGAGPTDV





Db 268708 GCGGGGCGGGCGGACCTTCCTGCTCCCGGCTTCGACCGGGCGCGGATCGAGC 268767  
QY 242 TGATCGAGCGGAGAGCTCGACCTTCCTGCGCGGCGTGCAGACCACTTACCACTCG 301  
Db 268768 TGATGCTGAAGAGGCGGCTCAACACCTTCACGCGGCTGCCGACCATGTTCTGCGCCCTCG 268827  
QY 302 TGGGGCGCGCGCGGAGCGGGTTTCTCGCGCGGAGCTGCGGATCGGCTGGCGGGG 361  
Db 268828 ---CCGCCGCGGCGCGCGGCGCGGCTGCCGAGCTGCGGGTGTGCATCTCCGGCG 268884  
QY 362 GCGCGGTCTCGGCGCGCGGCTGCCGAGCGAGTTCCGAGAGACCTTCGGGGTCCCGCTGA 421  
Db 268885 GTGCTCTGCTGCGGCTGCGGCTCTCGAACGGTTCCGAGAGGCTTCGGGGCGAGATCT 268944  
QY 422 TCGACGCCCTACGCGAGCACCGAGACCTTCGGGGCGGATCCCATGAACCGCGCGGCGG 481  
Db 268945 ACGAGGGCTACGGGCTGTGCGAGACCTCTCCGGCGGCGGCTGTGAACCAAGCCGGTGTTCG 269004  
QY 482 CCGCGGTGAGGGCTC 497  
Db 269005 GCGCCAGGCGCGGCAC 269020

Search completed: June 17, 2004, 01:40:37  
Job time : 2283.33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 22:08:13 ; Search time 285 Seconds

(without alignments)

7467.885 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325

Perfect score: 501

Sequence: 1 agttctaacgtcgggcac.....ccgggtcgagggctgctgc 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04.\*

1: Geneseqn1980s.\*

2: Geneseqn1990s.\*

3: Geneseqn2000s.\*

4: Geneseqn2001as.\*

5: Geneseqn2001bs.\*

6: Geneseqn2002s.\*

7: Geneseqn2003as.\*

8: Geneseqn2003bs.\*

9: Geneseqn2003cs.\*

10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	53789	AAV211187	AAV211187 Amycolato
2	302.6	60.4	14055	AAU611170	AAU611170 Actinosyn
3	302.6	60.4	82746	AAU61224	AAU61224 Actinosyn
4	243.8	48.7	2614	AAQ06631	AAQ06631 Gene conf
5	184.2	36.8	77536	AAU14651	AAU14651 Nucleotid
6	163.8	32.7	18660	AAU58472	AAU58472 Nucleotid
7	98.8	19.7	2520	AAH21138	AAH21138 Amycolato
8	91.8	18.3	1578	AAU25763	AAU25763 B.dhminut
9	87.8	17.5	1584	ACA23280	ACA23280 Prokaryot
10	85.4	17.0	2006	ADC68433	ADC68433 Lolium pe
11	85.4	17.0	2014	ADC68538	ADC68538 Lolium pe
12	85.2	17.0	1530	ABQ90146	ABQ90146 M. capsul
13	82.2	16.4	1992	ABK13572	ABK13572 Ryegrass
14	82.2	16.4	2038	ABK13573	ABK13573 Ryegrass
15	82.2	16.4	2073	ADC68539	ADC68539 S. arundi
16	80.6	16.1	2006	AAH42474	AAH42474 cDNA sequ
17	80.6	16.1	2122	AAU25197	AAU25197 Maize 4-c
18	78.8	15.7	1855	ADC68535	ADC68535 Lolium pe
19	78.4	15.6	110000	AAI99682_39	Continuation (40 o
20	78.4	15.6	110000	AAI99682_40	Continuation (41 o
21	78.4	15.6	110000	AAI99683_39	Continuation (40 o
22	77.4	15.4	2039	ADC68432	ADC68432 Lolium pe
23	77.4	15.4	2041	ADC68536	ADC68536 Lolium pe

24	77.2	15.4	2284	6	ABK13571	Abk13571 Ryegrass
25	76.6	15.3	30001	2	AAU61016	AAU61016 Total DNA
26	76.6	15.3	30001	2	AAU05110	Aax05110 S. aureof
27	76	15.2	7673	2	AAV58229	Aav58229 Omega-cyc
28	74	14.8	1386	6	ABK73860	Abk73860 Bacillus
29	74	14.8	90600	6	ABQ78872	Abq78872 S. roseos
30	72.6	14.5	1934	9	ADC68537	Adc68537 S. arundi
31	71.6	14.3	135638	7	ABX34289	Abx34289 S. atroool
32	71.4	14.3	1986	4	AAH42473	Aah42473 cDNA sequ
33	71	14.2	1970	2	AAU25198	Aax25198 Maize 4-c
34	70.6	14.1	47988	9	ADE86070	Ade86070 Streptomy
35	70.4	14.1	7191	3	AAU97538	Aaa97538 Streptomy
36	68.8	13.7	88421	6	AAU40781	Aal40781 8842int 9
37	67	13.4	1704	5	AAH65415	Aah65415 C glutami
38	67	13.4	1827	4	AAU67757	Aaf67757 Corynebac
39	67	13.4	34980	5	AAH68525	Aah68525 C glutami
40	66	13.2	2124	3	AAU55810	Aac55810 S. lavend
41	66	13.2	2124	9	ADE10229	Adel0229 S. lavend
42	66	13.2	53500	3	AAU55842	Aac55842 Complete
43	66	13.2	53500	9	ADE10261	Adel0261 S. lavend
44	66	13.2	110000	4	AAI99682_34	Continuation (35 o
45	66	13.2	110000	4	AAI99683_34	Continuation (35 o

## ALIGNMENTS

## RESULT 1

AAV211187

ID AAV211187 standard; DNA; 53789 BP.

XX AAV211187;

XX AAV211187;

DT 24-JUL-1998 (first entry)

DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.

XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;

XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;

XX polyketide synthase; actinomycete; ansamycin; ds.

XX Amycolatopsis mediterranei.

XX Amycolatopsis mediterranei.

XX Key Location/Qualifiers

FT CDS 1825..15543

FT CDS /tag= a

FT /label= ORF A

FT /product= "polyketide synthase"

FT 15550..30759

FT /tag= b

FT /label= ORF B

FT /product= "polyketide synthase"

FT 30895..36060

FT /tag= C

FT /label= ORF C

FT /product= "polyketide synthase"

FT 36259..41325

FT /tag= d

FT /label= ORF D

FT /product= "polyketide synthase"

FT 41373..51614

FT /tag= e

FT /label= ORF E

FT /product= "polyketide synthase"

FT 51713..5293

FT /tag= f

FT /label= ORF F

FT /product= "polyketide synthase"

XX WO9807868-A1.

XX 26-FEB-1998.

PD 18-AUG-1997;

PF 97WO-EF004495.

XX 20-AUG-1996; 96EP-00810551.  
XX (NOVS ) NOVARTIS AG.  
XX Schupp T, Toupet C, Engel N;  
XX WPI; 1998-169172/15.  
XX P-PSDB; AAW52845, AAW52846, AAW52847, AAW52848, AAW52849, AAW52850.  
XX  
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used to  
XX produce rifamycin and rifamycin analogues.  
XX Claim 4; Page 53-102; 205pp; English.  
XX  
XX The present sequence represents a Amycolatopsis mediterranei rifamycin  
XX synthesis gene cluster DNA fragment from the present invention. The DNA  
XX fragment comprises a DNA region involved directly or indirectly in the  
XX gene cluster responsible for rifamycin synthesis, including the adjacent  
XX DNA regions to the right and left which, by reason of their function in  
XX connection with rifamycin biosynthesis, qualify as constituents of this  
XX rifamycin gene cluster, and functional fragments, derivatives or  
XX constituents of these. The Amycolatopsis mediterranei rifamycin synthesis  
XX gene cluster DNA fragment can be used for producing rifamycin, rifamycin  
XX analogues or precursors. It can also be used for inactivating or  
XX modifying genes involved in ansamycin or rifamycin biosynthesis. The DNA  
XX can be used for constructing mutant actinomycetes strains from which the  
XX natural rifamycin or ansamycin biosynthesis gene cluster has been partly  
XX or completely deleted. The DNA fragment can be used for assembling a  
XX library of polyketide synthases, which can be used for assembling a  
XX library of polyketides. A hybridisation probe of the invention can be  
XX used for identifying DNA fragments involved in the biosynthesis of  
XX ansamycins  
XX  
XX Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 U; 0 Other;  
Query Match 100.0%; Score 501; DB 2; Length 53789;  
Best Local Similarity 100.0%; Pred. No. 5.2e-77;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGTTCTTACACGTGGGACGACCGGGCGGCCCAAGGGCGTCGTGTCCACGACGCGCAAC 60  
DB 1825 ATGTTCTTACACGTGGGACGACCGGGCGGCCCAAGGGCGTCGTGTCCACGACGCGCAAC 1884  
QY 61 TGCCTCTGTGTCGTGCTTCTGCTAGTGGCGTTCCTCCCGGTTCGACACGACCGG 120  
DB 1885 TGCCTCTGTGTCGTGCTTCTGCTAGTGGCGTTCCTCCCGGTTCGACACGACCGG 1944  
QY 121 GTGCTCTGGCGGCTCCGCTGTTCCACAGCCTTTCCGACATCGCTCGCTGCTGCGCC 180  
DB 1945 GTGCTCTGGCGGCTCCGCTGTTCCACAGCCTTTCCGACATCGCTCGCTGCTGCGCC 2004  
QY 181 ACCGTGTCTGGGGCCAGCGTCGGATCCGACGCGACGCGCTCCGCGACGAGTATCGG 240  
DB 2005 ACCGTGTCTGGGGCCAGCGTCGGATCCGACGCGACGCTCCGCGACGAGTATCGG 2064  
QY 241 CTGATCTGAGGCGGAGAGCTGACCTTCTGCTGCGCGGCTGCGACACCTACCACTG 300  
DB 2065 CTGATCTGAGGCGGAGAGCTGACCTTCTGCTGCGCGGCTGCGACACCTACCACTG 2124  
QY 301 GTGCGGCGCGCGCGCGCGGTTCTTCGCGCGCGGCTTCGAGTCCGCGCTGCGCGG 360  
DB 2125 GTGCGGCGCGCGCGCGCGGTTCTTCGCGCGCGGCTTCGAGTCCGCGCTGCGCGG 2184  
QY 361 GCGCGGCTCTCGGCGCGCGGCTGCGAGAGCGAGTTCGAAGAGACCTTCGCGGTTCCGCTG 420  
DB 2185 GCGCGGCTCTCGGCGCGCGGCTGCGAGAGCGAGTTCGAAGAGACCTTCGCGGTTCCGCTG 2244  
QY 421 ATCGACGCTTACGCGGACGACCGAGACCTTCGCGGCGGATTCACATGAAACCGCGGACGGC 480  
DB 2245 ATCGACGCTTACGCGGACGACCGAGACCTTCGCGGCGGATTCACATGAAACCGCGGACGGC 2304  
QY 481 GCCCGCGTCGAGGGTCTCGTGC 501

DB 2305 GCCCGCGTCGAGGCTCGTGC 2325  
RESULT 2  
ID AAL61170 standard; DNA; 14055 BP.  
XX AAL61170;  
XX AAL61170;  
XX 22-SEP-2003 (first entry)  
XX Actinosynnema pretiosum polyketide synthase (PKS) gene #1.  
XX Actinosynnema pretiosum polyketide synthase; enzyme; PKS;  
XX Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;  
XX Gene; ds.  
XX Actinosynnema pretiosum.  
XX WO2003045312-A2.  
XX 05-JUN-2003.  
XX 21-NOV-2002; 2002WO-US037547.  
XX 21-NOV-2001; 2001US-0332158P.  
XX (UNIW ) UNIV WASHINGTON.  
XX Floss HG, Yu T, Leistner E;  
XX WPI; 2003-493374/46.  
XX Novel maytansinoid produced by bacterial host cell transformed with  
XX expression vector comprising open reading frame from ansamitocin gene  
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX Disclosure; Page 45-53; 160pp; English.  
XX The invention relates to maytansinoid produced by bacterial host cell  
XX transformed with expression vector comprising open reading frame from  
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
XX useful as a potent antitumor agent. The present sequence is A. pretiosum  
XX ansamitocin gene cluster I polyketide synthase (PKS) gene  
XX  
XX Sequence 14055 BP; 1304 A; 5585 C; 5518 G; 1648 T; 0 U; 0 Other;  
Query Match 60.4%; Score 302.6; DB 7; Length 14055;  
Best Local Similarity 75.2%; Pred. No. 3.5e-43;  
Matches 377; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 1 ATGTTCTTACACGTGGGACGACCGGGCGGCCCAAGGGCGTCGTGTCCACGACGCGCAAC 60  
DB 484 ATGCTCTTACACCTTCCGGACACCGGGCGGCCCAAGGGCGTCGTGTCCACGACGCGCAAC 543  
QY 61 TGCCTCTGTGTCGTGCTTCTGCTAGTGGCGTTCCTCCCGGTTCGACACGAGACCGG 120  
DB 544 TGCCTCTGTGTCGTGTCGCGCTGCTAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 603  
QY 121 GTGCTCTGGCGGCTCCGCTGTTCCACAGCCTTTCCGACATCGCTCGCTGCTGCTGCTGCTG 180  
DB 604 GTGCTCTGGCGGCTCCGCTGTTCCACAGCCTTTCCGACATCGCTCGCTGCTGCTGCTGCTG 663  
QY 181 ACCGTGTCTGGGGCCAGCGTCGGATCCGCGACGCGGCTCCGCGACGAGTATCGG 240  
DB 664 ACCCGCTCGCGCGGCGGACCGCGCTGCTGCGAGCGGCTCCCGCGGCTGCTGCTGCTGCTG 723  
QY 241 CTGATCTGAGGCGGAGAGCTGACCTTCTGCGCGCGGCTTCGACGAGGCTGCTGCTGCTGCTG 300  
DB 724 GCGCTGACGAGGAGCGCTCCACGCTGCTCGCGGCGGCTGCGCGGCTGCGCGGCTGCTG 783  
QY 301 GTGCGGCGCGCGCGCGGCTTCTTCGCGCGCGGCTTCTTCGCGCGGCTGCGGCTGCGCGG 360



Db 784 GTGCGCGCGCGGTGAGCGCGGTTTCTCGCGCCGACCTGCGGGTCGGGCTGGTCCGC 843  
Qy 361 GCGCGGTCTCTCGCGCGCGGTGCGAGCGAGTTGAGAGACCTTCGGGGTCCGCTG 420  
Db 844 GCGCGCATACCAACGCGCGGAGCTGCGGCGCGCCGTCGAGCAGCGGTTCGGCGTCCGCTC 903  
Qy 421 ATCGACGCTTACCGCGCAGCAGCGGAGACCTGCGGCGCGATCACATGAACCCGCGGACGGC 480  
Db 904 ATCGACGCTTACCGCGCAGCAGCGGAGAGCTGCGGTCGATCACCGTGAACCTGCGCGACGGG 963  
Qy 481 GCGCGGTGAGGGCTGCTGC 501  
Db 964 CCAGGCTCAGGGCTCCAGC 984

RESULT 3  
AAL61224/c  
ID AAL61224 standard; DNA; 82746 BP.  
XX AC AAL61224;  
XX DT 22-SEP-2003 (first entry)  
XX DE Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.  
XX KW Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX OS Actinosynnema pretiosum.  
XX PN WO2003045312-A2.  
XX PD 05-JUN-2003.  
XX PF 21-NOV-2002; 2002WO-US037547.  
XX PR 21-NOV-2001; 2001US-0332158P.  
XX PA (UNIW ) UNIV WASHINGTON.  
XX PI Floss HG, Yu T, Leistner E;  
XX WPI; 2003-493374/46.  
XX Novel maytansinoid produced by bacterial host cell transformed with  
PT expression vector comprising open reading frame from ansamitocin gene  
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX Claim 7; Page 105-152; 160pp; English.  
XX The invention relates to maytansinoid produced by bacterial host cell  
CC transformed with expression vector comprising open reading frame from  
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
CC useful as a potent antitumour agent. The present sequence is A. pretiosum  
CC ansamitocin ansamitocin biosynthetic gene cluster I  
XX  
SQ Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;  
Query Match 60.4%; Score 302.6; DB 7; Length 82746;  
Best Local Similarity 75.2%; Pred. No. 2.9e-43;  
Matches 377; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
Qy 1 ATGTTCTACACGTGCGGCACAGCGCGCGGCCCAAGGCGTGTCTCCAGCGCGCAAC 60  
Db 36558 ATGCTTACACCTTCGGCAGCAGCGCGCGGCCCAAGGCGTGTCTCTCCAGCGCGCAAC 36499  
Qy 61 TGCTCTGGTCCGTGCTTCTGCTAGTGCCTGTCCTCCCGGTTGTGGACCGGACCGG 120  
Db 36498 TGCTCTGGTCCGTGCGCGCTGCTAGCGCGCGTGTCTGCTGACCCCGGACCGG 36439  
Qy 121 GTGCTTGGCGCGTCCCGCTGTTCACAGCGCTTTCGACATCGCTCGCTGTCGCGC 180  
Db 36438 GTGCTTGGCGCGTCCCGCTGTTCCACAGCGCTTTCGACATCGCGCGGTGCTGGCCACC 36379

Qy 181 ACCGTGTGCGGGCCAGCGTCCGATCGCGACGCGAGCTCCGCGACGACGTGATCGG 240  
Db 36378 ACCGCGTTCGGCGCCACACCCCGGTGTCGAGCGGCGCTCCCGCAGAGCTGCTGGCC 36319  
Qy 241 CTGATCGAGCGGAGAGCTCGACCTTCTCTGGCCGCGGTGCGGACCACTACCACTG 300  
Db 36318 GCCTCTGACGAGGAGCGCTCCACGGTGTCTCGCGGCGTGCACGACTACCACTG 36259  
Qy 301 GTGCGGGCGCGCGGAGCGCGGTTCCTCGCGCGGAGCTGCGGATCGGCTGCGCGGG 360  
Db 36258 GTGCGCGCGCGCGGTGAGCGCGGTTCCTCGCGCGGAGCTGCGGCTGCGGCTG 36199  
Qy 361 GCGCGGTCTCTCGCGCGCGCGCTGCGAAGCGAGTTCGAAAGAGACCTTCGGGGTCCCGCTG 420  
Db 36198 GCGCGATCACCAACGCGCGGAGCTGCGGCGCGCTCGAGCAGCGTTCGCGGTGCGCTC 36139  
Qy 421 ATCGACGCTTACCGCGCAGCAGCGGAGCTGCGGCGGATCACATGAACCCGCGGACGGC 480  
Db 36138 ATCGACGCTTACCGCGCAGCAGCGGAGCTGCGGCGGATCACATCGCTGACCTGCGGCGG 36079  
Qy 481 GCGCGGTGCGAGGGCTGCTGC 501  
Db 36078 CCAGGCTCAGGGCTCCAGC 36058

RESULT 4  
AAQ06631  
ID AAQ06631 standard; DNA; 2614 BP.  
XX AC AAQ06631;  
XX DT 24-OCT-2003 (revised)  
XX DT 27-AUG-2003 (revised)  
XX DT 25-FEB-1991 (first entry)  
XX DE Gene conferring teicoplanin resistance.  
XX KW Dalbaheptide antibiotic resistance; probe; ss.  
XX OS Actinoplanes teichomyceticus; strain ATCC 31121.  
XX PN EP39328-A.  
XX DT 28-NOV-1990.  
XX PF 14-MAY-1990; 90EP-00109074.  
XX PR 23-MAY-1989; 89EP-00109273.  
XX (LEPE ) GRUPPO LEPETIT SPA.  
XX PI Denaro M, Lorenzetti R, Moroni MA, Sosio MA;  
XX WPI; 1990-356009/48.  
XX DNA sequence conferring specified antibiotic resistance - obtd. as DNA  
PT fragment by Bam HI digestion of Actinoplanes teichomyceticus strain.  
XX Claim 6; Page 11; 16pp; English.  
XX The total genomic DNA of A. teichomyceticus ATCC 31121 was digested and  
CC ligated with pJ3702 (from S. lividans). The ligation mixt. was used to  
CC transform S. lividans 66 protoplast. One clone was selected for its  
CC ability to grow in the presence of 20 mg/l teicoplanin. The plasmid in  
CC this clone was designated pTR168 and was found to contain the sequence  
CC below. The teico-R conferring sequence can be used as a marker in DNA  
CC manipulations. It can also be introduced into dalbaheptide antibiotic  
CC producing strains to improve anti-biotic prodn. yields. It can also be  
CC used as a probe to detect corresp. regions in the genomes of other  
CC producers of dalbaheptide antibiotics. (Updated on 27-AUG-2003 to correct  
XX OS field.) (Updated on 24-OCT-2003 to standardise OS field)  
SQ Sequence 2614 BP; 310 A; 984 C; 994 G; 326 T; 0 U; 0 Other;





DT 31-OCT-2000 (first entry)  
 XX Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 31-40.  
 DE BLM gene cluster; bleomycin gene cluster; polyketide metabolite;  
 XX bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;  
 KW thiazoline; bithiazoline; microbial metabolite; sugar; ss.  
 XX Streptomyces verticillus.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FT 1. .354  
 CDS /tag= a  
 FT /transl\_except= (pos: 1. .3, aa: Met)  
 FT /note= "ORF31; encodes AAB07580"  
 CDS 925. .9450  
 FT /tag= b  
 FT /note= "ORF32; encodes AAB07581"  
 CDS 9447. .10802  
 FT /tag= c  
 FT /note= "ORF33; encodes AAB07582"  
 CDS 10864. .11877  
 FT /tag= d  
 FT /note= "ORF34; encodes AAB07583"  
 CDS 11888. .12640  
 FT /tag= e  
 FT /note= "ORF35; encodes AAB07584"  
 CDS 12637. .13920  
 FT /tag= f  
 FT /note= "ORF36; encodes AAB07585"  
 CDS 13928. .14746  
 FT /tag= g  
 FT /transl\_except= (pos: 1. .3, aa: Met)  
 FT /note= "ORF37; encodes AAB07586"  
 CDS 14743. .16479  
 FT /tag= h  
 FT /note= "ORF38; encodes AAB07587"  
 CDS 16476. .17462  
 FT /tag= i  
 FT /transl\_except= (pos: 1. .3, aa: Met)  
 FT /note= "ORF39; encodes AAB07588"  
 CDS 17846. .18659  
 FT /tag= j  
 FT /note= "ORF40; encodes AAB07589; no termination codon given"  
 XX WO200040704-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 XX 06-JAN-2000; 2000WO-US000445.  
 XX  
 XX 06-JAN-1999; 99US-0115435P.  
 PR  
 XX 05-FEB-1999; 99US-0118848P.  
 PR  
 XX 05-JAN-2000; 2000US-00477962.  
 XX  
 XX (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Shen B, Du L, Sanchez C, Chen M, Edwards DJ;  
 PI  
 XX WPI; 2000-465974/40.  
 DR  
 DR P-PSDB; AAB07580, AAB07581, AAB07582, AAB07583, AAB07584, AAB07585,  
 DR AAB07586, AAB07587, AAB07588, AAB07589.  
 XX  
 XX New bleomycin gene cluster components useful for peptide and/or  
 PT polyketide metabolites, especially bleomycin, production and for  
 FT chemically modifying biological molecules.  
 FT  
 XX  
 XX Claim 8; Page 137-153; 162pp; English.  
 PS  
 XX  
 XX The present sequence represents the BLM (Bleomycin) gene cluster,  
 CC containing open reading frames (ORFs) 31-40. The proteins encoded by the  
 CC gene cluster are useful for producing peptides and/or polyketide

CC metabolites, especially bleomycin or bleomycin analogues. They are also  
 CC useful for chemically modifying biological molecules to produce branched  
 CC methyl groups, and for coupling amino acids and fatty acids. They may be  
 CC reacted with an apo-carrier protein and coenzyme A to produce a holo-  
 CC carrier protein. The BLM gene cluster or catalytic domains can be used  
 CC individually or collectively to produce thiazolidine, thiazoline,  
 CC bithiazoline and bithiazoline-containing microbial metabolites. The BLM  
 CC gene cluster may also be used to produce sugars  
 XX  
 XX Sequence 18660 BP; 2365 A; 7239 C; 6510 G; 2546 T; 0 U; 0 Other;  
 SQ  
 Query Match 32.7%; Score 163.8; DB 3; Length 18660;  
 Best Local Similarity 61.0%; Pred. No. 1.4e-19;  
 Matches 303; Conservative 0; Mismatches 187; Indels 7; Gaps 2;  
 QY 1 ATGTTCTACACGTCGGGCGACGCGGCGGCCCAAGGGCGTCTGTCCACGACGACGCAAC 60  
 DB 18164 ATCTCTTACACGTCGGGCGACGCGGCGGCGGTACGTGCGGTCTGGGGCGCAGGACCGG 18223  
 QY 61 TGGCTCTGTGTCGCTTCTCTCTAGTCGCTTCCCGGGTTTGTCCGACGACGACCGG 120  
 DB 18224 GCGCTGTGTCGCTTCTGGGGCGGCGGTACGTGCGGTCTGGGGCGCAGGACCGG 18283  
 QY 121 GTGCTCTGGGCGCTCCGCGTTCACAGGCTTTCCACATCGCCTGCTGTCGCGCC 180  
 DB 18284 CTGTTGTGGCGCTGCCCATGTTCCACGCTACGCGCACTCGCTGTGCTGCGGGTG 18343  
 QY 181 ACCGTGTGCGGGCGCAGCGTCGCGATCGCGACGCGGCGGTCCGCGACGACGTATCGCG 240  
 DB 18344 GTGCGCTGCGGCGGAGCGGTACCTCTCGACCGGGGCGGAGCGTGG-----TCCGG 18397  
 QY 241 CTGATCGAGCGCGAGAGCTCGACCTTCTGCGCGGGTTCGCGACCACTACCAACCTG 300  
 DB 18398 GCGCTTGAGGAACAGCGGTGACGCGTCTGCGCGGTACCGCCACTACCGCTGCTC 18457  
 QY 301 GTGGGGCGCGCGCGAGCGCGGTTCCTCGCGCGGAGCTTCGCGATCGGCTGGCGCGG 360  
 DB 18458 ACGAGCGCTTCCGCGACGCGCCCGCGCCACCGCGCGGCTGCGACTGTGCTACCGGG 18517  
 QY 361 GCGCGCTCTCTCGCT 419  
 DB 18518 GGCTGCGCGCTGCT 18577  
 QY 420 GATCGACGCTTACGCGACGACCGGAGCTTCGCGGGCGGATCACCATGAACCGCGGACGG 479  
 DB 18578 GCTCGAGGTTACGCGAGTACCGGAGCTTCGCGCAAGATCAGCGCTTACGCGGCTCGCGG 18637  
 QY 480 CGCCCGCGTCGAGGGCT 496  
 DB 18638 CTCGCGGAGGCGGTT 18654  
 RESULT 7  
 AAH21138  
 ID AAH21138 standard; cDNA; 2520 BP.  
 XX  
 XX AAH21138;  
 AC  
 XX  
 XX 11-SEP-2001 (first entry)  
 DT  
 XX Amycolatopsis sp HR167 cDNA encoding ech and fcs proteins.  
 DE  
 XX Vanillin; ech; enoyl-CoA hydratase/aldolase; fcs; ferulic acid;  
 KW feruloyl-CoA synthetase; ss.  
 XX  
 XX Amycolatopsis sp.  
 OS  
 XX Key Location/Qualifiers  
 FT 114. .117  
 FT RBS /tag= a  
 FT CDS 124. .984  
 FT /tag= b  
 FT /product= "ech"

RBS 978. .981  
 CDS /\*tag= c  
 987. .2462  
 /\*tag= d  
 /product= "fcs"  
 DE19960106-A1.  
 21-JUN-2001.  
 14-DEC-1999; 99DE-01060106.  
 14-DEC-1999; 99DE-01060106.  
 (HAAR ) HAARMANN & REIMER GMBH.  
 Rabenhorst J, Steinbuechel A, Priefert H, Achterholt S;  
 WPI; 2001-399437/43.  
 P-PSDB; AAB86276, AAB86277.  
 New enzymes from Amycolatopsis, useful for synthesis of vanillin from  
 ferulic acid, also nucleic acid encoding them and recombinant host cells.  
 Claim 8; Page 9-14; 18pp; German.  
 This invention describes novel enzymes (I) from Amycolatopsis sp HR167  
 involved in the synthesis of vanillin from ferulic acid (FA). This  
 sequence represents a cDNA fragment encoding the enoyl-CoA  
 hydratase/aldolase (ech) and the feruloyl CoA-synthetase (fcs) which are  
 described in the method of the invention  
 Sequence 2520 BP; 355 A; 895 C; 929 G; 341 T; 0 U; 0 Other;  
 Query Match 19.7%; Score 98.8; DB 4; Length 2520;  
 Best Local Similarity 54.5%; Pred. No. 2.1e-08;  
 Matches 243; Conservative 0; Mismatches 197; Indels 6; Gaps 2;  
 2 TGTCTACAGTCGGGCAAGACCGGGGGCCCAAGGGGTCGTGTCACGAGCGCAACT 61  
 1438 TGATGTACACTCGGGCAGCAGCGGACGCCCAAGGGCGCATGTCAACCCACGCGCAAC 1497  
 62 GCCTTGCTCGTCGCTTCTGCTACGTCGCTTCCCGGGTGTTCGACACGACGCGG 121  
 1498 TCACCTGGAACTGGCTCAACGTCTCTGTGTGAGA---CCGACCTGGCGAGCGAGCGGG 1554  
 122 TGCTTGGCCGCTCCCGCTGTTCCACAGCCCTTTGCAATCGCCTGCTGCTCCGCCA 181  
 1555 CACTGGTCGCGCGCGCTGTTCCAGCGCGCGGCTCGGCATGGTGTGCTGCCACCC 1614  
 182 CCGTGGTCGGGGCCAGCGTCGCGATCGCGGATCGCGCGAGCGAGCTCGCGCGATGCGGC 241  
 1615 TGCTCAAGGGGGCAGCGGTGATCTCTGCACTCGCGCTTCGACCCCGCGCGCTGCTCG 1674  
 242 TGATCGAGCGGAGCTCGACCTTCTGCGCGGCTGCGACCACTTACCACTCGG 301  
 1675 CGGTGGAAAGAGAGCGGTTCAGCTCGGTTCGCGGTGCCACGATGTACAGGCGATCG 1734  
 302 TCGCGCGCGCGCGCAGCGCGGTTCCTCGCGCGCGAGCTCGCGATCGCGCTGGCGGG 361  
 1735 CGCGCACCGCGGTGTGCGCAGCGCGGACCTCTCCAGCTCGGACCTGCTGTGCGGG 1794  
 362 GCGCGGTCTCTCGCGCGCGGGCTCGGAAGAGAGTTCGAAGAGACCTTCGCGGGTCCGCTGA 421  
 1795 GCGCGCGGTGCGCGCGGACCTCGCGAGCGGTACCTCGA---CGCGGGCTCGGTTG 1851  
 422 TCGACGCTTACGCGAGCAGCGGAC 447  
 1852 TGCAGGGCTACGCGCATGACCGAGGCC 1877

AAX25763 standard; DNA; 1578 BP.  
 AC AAX25763;  
 08-JUN-1999 (first entry)  
 B.diminuta pimelyl CoA synthase coding sequence.  
 Pimelyl CoA synthase; pimelyl CoA; pimelic acid; coenzyme A; ds.  
 Brevundimonas diminuta.  
 JP11046763-A.  
 23-FEB-1999.  
 31-JUL-1997; 97JP-00205791.  
 31-JUL-1997; 97JP-00205791.  
 (MITU ) MITSUBISHI CHEM CORP.  
 WPI; 1999-208106/19.  
 P-PSDB; AAW99453.  
 New pimelyl CoA synthase gene - useful for preparation of pimelyl CoA  
 from pimelic acid and coenzyme A.  
 Claim 2; Page 8-10; 11pp; Japanese.  
 This sequence represents the coding region for the pimelyl CoA synthase  
 protein from Brevundimonas diminuta strain IF03149. The sequence can be  
 used for the preparation of pimelyl CoA from pimelic acid and coenzyme A  
 Sequence 1578 BP; 273 A; 502 C; 533 G; 270 T; 0 U; 0 Other;  
 Query Match 18.3%; Score 91.8; DB 2; Length 1578;  
 Best Local Similarity 51.2%; Pred. No. 3.4e-07;  
 Matches 241; Conservative 0; Mismatches 227; Indels 3; Gaps 1;  
 1 ATGTTCTACAGTCGGGCAAGACCGGGGGCCCAAGGGCGTCGTGTCACGAGCGCAAC 60  
 535 ATCTGTTCACACAGCGGTACGACCGGAACCCGAGGGCGGGTGGGTAGCCACTGTGCG 594  
 61 TGCCTCTGTCGCTCGCTTCTGCTACGTGCGCTTCCCGGGTGTTCGGAACAGGACCGG 120  
 595 ATTGTGAGCAGGTGCATCGGGGCTCATTGACGGGGTTCAAGTGTGGAACGAGACGCTG 654  
 121 GTGCTGTGGCGCTCCCGCTGTTCACAGCCTTTCGACATCGCTGCTGCTGTCGCGC 180  
 655 ACGTTTGGCGGGCTGCCGCTGTTCATACGTTCGCGCCAGACTTCGGTCTGAAACATGGCG 714  
 181 ACCGTGTTCGGGGCCAGGTCCGGATCGCCGAGCGCAGCTCCGCGCAGCAGTGTGCGG 240  
 715 TTCGCGCGCGGCGCGCTCATCTGATGCCACGATTTGACGGGCAACAGGCTCTCGAA 774  
 241 CTGATCGAGGCGGAGAGCTTCGACCTTCTCGCGCGCTGCGGACCACTTACCACTCG 300  
 775 CTGATGTGAACACCAAGCAACGGTGTTCGCGGGTCCGACCATGTACGCGCTC 834  
 301 GTGCGGGCGCGCGCGCGGTTCCTCGCGCGCGAGCTTCGGATCGGCTTGGCGCGG 360  
 835 GTGAGGCGCGCG---CGGACCGACGCGCGCGCGCTTCGGTTCGCGAATTCGCGC 891  
 361 GCGCGGTCTCTCGCGCGCGGTGCGAGCGAGTTGAGAGACCTTCGGGTTCGCGCTG 420  
 892 GGTGCGCGCTTCCGCTGCTGCTGAGGGCTTTGAGAGGCGCTTCGGGTGCGCGC 951  
 421 ATCGAGCGCTACGCGACGACCGAGACCTTCGCGGGCGGATCACCATGAACCG 471  
 952 CACGAGGGCTACGGGCTTACCAGACAGCGCCCGACGGTCAACCGTCAATACG 1002





P-PSDB; ADC68468.

New polynucleotide encoding polypeptides from Loliium perenne or Festuca arundinacea, useful for modulating the biosynthesis of lignin, fructan or tannin in a plant.

Claim 1; SEQ ID NO 143; 240pp; English.

The present invention describes isolated polynucleotides (I) encoding proteins (II) from Loliium perenne and Festuca arundinacea which are active in lignin, fructan and tannin biosynthetic pathways. Also described: (1) an isolated oligonucleotide probe or primer comprising at least 10 contiguous residues complementary to 10 contiguous residues of (I); (2) a kit comprising the oligonucleotide probe or primer; (3) a genetic construct comprising (I); (4) a transgenic plant cell comprising the genetic construct of (3); (5) a plant or its seed, fruit or progeny comprising the transgenic plant cell of (4); (6) modulating one or more of the lignin, fructan or tannin compositions of a plant; (7) producing a plant having one or more of the lignin, fructan or tannin compositions; and (8) modifying the activity of (II) involved in a lignin, fructan or tannin biosynthetic pathway in a plant. (I) can be used for modulating the biosynthesis of lignin, fructan or tannin in a plant. The present sequence is used in the exemplification of the present invention.

Sequence 2006 BP; 405 A; 630 C; 615 G; 356 T; 0 U; 0 Other;

Query Match 17.0%; Score 85.4; DB 9; Length 2006;  
Best Local Similarity 51.5%; Pred. No. 4.1e-06;  
Matches 253; Conservative 0; Mismatches 226; Indels 12; Gaps 2;

XX QY 2 TGTTCTACAGTCGGGCGCAGCCGGGGGCCAAGGGCGTGTGTCACGAGCGCAACT 61  
XX Db 706 TGCCCTACTCTCCGCGCACCCAGCGGGTCCCACAAGGGCGTCATGCTCACCCACCGCAGCC 765

XX QY 62 GCCTTGCTGGTCTGC-----TTCTCTACGTGCGCTTCCCCGGTGTGCGACC 112  
XX Db 766 TGGTCACAGCGTCCGCCAGCTGGTTGACGGTTCGAACAAAGTGTGTTCAACAAGG 825

XX QY 113 AGAACCGGGTGTCTGGCGGCTCCCGCTGTTCCACAGCCTTTCCGACATCGCTGCGTCC 172  
XX Db 826 ACAGCGCGTGTGTGCTGCTGCTGCGCTGTTCCCATCTACTCGTGCACAGGTGTGTC 885

XX QY 173 TGTCCGCCACCGTGTGCGGGCCAGCGTCCGGATCGCGACGCGAGTTCGCCCGCAGCAG 232  
XX Db 886 TGGCGGGGCTCCGCTCGCGCGCGCCATCGTTCATCATCGCAAGTTCGACGTCGCGCGC 945

XX QY 233 TGATGGGCTGATCGAGCGGAGACTCGACTTCTGCGCGGCTGCGCACCACTTACC 292  
XX Db 946 TGGTGGACCTCTCGCGCGCCACCGATCATCATCGCGCCATTGTCGCGCCCATCTGTG 1005

XX QY 293 ACCACTGTGCGGGCGCGCCCGGAGCGGTTTCTCGCGCGCGAGCTCGGGATCGGCC 352  
XX Db 1006 TGGAGATCGCCAGAGCGACCGCGTTCGCGCGCGACGACTCGCATCCATCGCATGTG 1065

XX QY 353 TGGCCCGGGCGCGGTCTTCGCGCGCGGCTCGGAAGCGAGTTC---GAAAGACCTTCG 409  
XX Db 1066 TCTCCGCGCGCGGCCCATGCGGCAAGGACCTCGAGAGCGCTTTCATGCGCAAGATCCCA 1125

XX QY 410 GGGTCCCGCTGATCGAGCGCTAGCGCAGCACCAGACCTTCGGGGCGGATCCCATGACC 469  
XX Db 1126 ACGCCGTGTCTCGACAGGGTTACGGGATGACTGAGGCGCGGCGCGTGTGCGCATGTG 1185

XX QY 470 CGCGCGGACGAC 480  
XX Db 1186 TGGCGTTCGCC 1196

RESULT 11  
ADC68538  
ID ADC68538 standard; cDNA; 2014 BP.  
XX AC  
XX AC  
XX AC

18-DEC-2003 (first entry)  
Loliium perenne lignin/tannin biosynthesis protein cDNA SEQ ID NO:30.  
Loliium perenne; Festuca arundinacea; lignin; fructan; tannin;  
biosynthetic pathway; plant; gene; ss.  
Loliium perenne.  
WO2003040306-A2.  
15-MAY-2003.  
07-NOV-2002; 2002WO-NZ000239.  
07-NOV-2001; 2001US-0337703P.  
(GENE-) GENESIS RES & DEV CORP LTD.  
(WRIG-) WRIGHTSON SEEDS LTD.  
Demmer J, Forster RL, Gibson JB, Shenk MA, Norriss MG, Glenn M;  
Saulsbury KM, Hall C;  
WPI: 2003-441544/41.  
P-PSDB; ADC68382.  
New polynucleotide encoding polypeptides from Loliium perenne or Festuca arundinacea, useful for modulating the biosynthesis of lignin, fructan or tannin in a plant.  
Claim 1; SEQ ID NO 30; 240pp; English.  
The present invention describes isolated polynucleotides (I) encoding proteins (II) from Loliium perenne and Festuca arundinacea which are active in lignin, fructan and tannin biosynthetic pathways. Also described: (1) an isolated oligonucleotide probe or primer comprising at least 10 contiguous residues complementary to 10 contiguous residues of (I); (2) a kit comprising the oligonucleotide probe or primer; (3) a genetic construct comprising (I); (4) a transgenic plant cell comprising the genetic construct of (3); (5) a plant or its seed, fruit or progeny comprising the transgenic plant cell of (4); (6) modulating one or more of the lignin, fructan or tannin compositions of a plant; (7) producing a plant having one or more of the lignin, fructan or tannin compositions; and (8) modifying the activity of (II) involved in a lignin, fructan or tannin biosynthetic pathway in a plant. (I) can be used for modulating the biosynthesis of lignin, fructan or tannin in a plant. The present sequence is used in the exemplification of the present invention.  
Sequence 2014 BP; 413 A; 628 C; 618 G; 355 T; 0 U; 0 Other;

Query Match 17.0%; Score 85.4; DB 9; Length 2014;  
Best Local Similarity 51.5%; Pred. No. 4.1e-06;  
Matches 253; Conservative 0; Mismatches 226; Indels 12; Gaps 2;

QY 2 TGTTCTACAGTCGGGCGCAGCCGGGGGCCAAGGGCGTGTGTCACGAGCGCAACT 61  
Db 707 TGCCCTACTCTCCGCGCACCCAGCGGGTCCCACAAGGGCGTCATGCTCACCCACCGCAGCC 766

QY 62 GCCTCTGGTCTGCTGCG-----TTCTCTACGTGCTGCTTCCCGGGTGTGCGACC 112  
Db 767 TCGTCACAGCGTTCGCCAGCTGGTTGACGGGTTCGAACCCCAAACGTGTGCTTCAACAAGG 826

QY 113 AGAACCGGGTGTCTGCGCGCTCCCGCTGTTCCACAGCCTTTCCGACATCGCTGCGTCC 172  
Db 827 ACGACGCGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 886

QY 173 TGTCGCCACCGTGTGCGGGCCAGCGTCCGATCGCGAGCTCCGCGCGAGCTCCGCGCGAGC 232  
Db 887 TGGCGGGGCTTCGCGTGTGCGCGCGCCCATCTGTCATGTCGCGAAGTTCGACGTTCGGCGCG 946

QY 233 TGATGCGGCTGATCGAGGGGAGAGCTCGACCTTCTGCGCGCGGCTGCGCGACCACTACC 292  
Db 947 TGGTGGACCTCGTTCGCGCGCGCACCGCATCACCATCGCGCCATCGTGTGCGCCCATCGTG 1001



Thu Jun 17 08:25:02 2004

XX	PA	(VICT-) STATE VICTORIA DEPT NATURAL RES & ENVIRO.	
XX	PA	(UYAD-) UNIV ADELAIDE.	
XX	PA	(ITMA-) INT MAIZE & WHEAT IMPROVEMENT CENT.	
XX	PA	(SAUS-) STATE SOUTH AUSTRALIA SOUTH AUSTRALIAN R.	
XX	PA	(UYSC-) UNIV SOUTHERN CROSS.	
XX	PA	(DAIR-) DAIRY RES & DEV CORP.	
XX	PI	Spangenberg GC, Lidgett AJ, Heath RL, McInnes RL, Lynch DP;	
XX	PI	WPI; 2002-097993/13.	
XX	DR	P-PSDB; AAU75085.	
XX	PT	Novel nucleic acid encoding enzymes involved in lignin biosynthetic	
XX	PT	pathway from ryegrass or fescue species useful for modifying lignin	
XX	PT	biosynthesis in plants and as a molecular genetic marker.	
XX	PS	Claim 3; Fig 3; 148pp; English.	
XX	XX	This invention represents purified or isolated nucleic acid and protein	
XX	XX	sequences of enzymes involved in lignin biosynthesis. The enzymes of the	
XX	XX	invention are 4 coumarate CoA-ligase (4CL), cinnamoyl-CoA reductase (CCR)	
XX	XX	and cinnamyl alcohol dehydrogenase (CAD) from a ryegrass (Lolium sp.) or	
XX	XX	fescue (Festuca sp.). The invention also comprises an isolated regulatory	
XX	XX	element from the nucleic acid sequences and a plant cell or seed	
XX	XX	transformed with the nucleic acid. An isolated regulatory element from	
XX	XX	these nucleotide molecules is useful for expressing an exogenous gene in	
XX	XX	plant cells. The nucleotide sequences of the invention and vectors	
XX	XX	containing these sequences are useful for modifying lignin biosynthesis	
XX	XX	in a plant and are useful as a molecular genetic marker for qualitative	
XX	XX	trait loci (QTL) tagging, QTL mapping, DNA fingerprinting and in marker	
XX	XX	assisted selection, in forage and turf grass improvement, e.g. tagging	
XX	XX	QTLs for dry matter digestibility, herbage quality, drought tolerance,	
XX	XX	regrowth after cutting and grazing, cold tolerance, drought tolerance,	
XX	XX	tiller survival and plant persistence. The present sequence represents	
XX	XX	the perennial ryegrass 4-coumarate Co-A-ligase 2 (Lp4CL2) cDNA of the	
XX	XX	invention	
XX	XX	Sequence 1992 BP; 409 A; 622 C; 576 G; 385 T; 0 U; 0 Other;	
XX	XX	Query Match 16.4%; Score 82.2; DB 6; Length 1992;	
XX	XX	Best Local Similarity 51.1%; Pred. No. 1.4e-05;	
XX	XX	Matches 251; Conservative 0; Mismatches 228; Indels 12; Gaps 2;	
QY	2	TGTTCTACAGTCGGGACACACCGGGCGCCCAAGGCGTCGTGTCACACGACGGCACT 61	
DB	636	TGCCGTACTCGTCGGGCACACCGGGCGTCGCCAAGGCGTCATGCTCACCCACCGCAGCC 695	
QY	62	GCCTCTGGTCCGTGCTTC-----CTGCTACGTGCGGTTCCCGGGTTGTGGACC 112	
DB	696	TCGTACACACGCTGCGCCACAGAGTGACGGTGAGAACCCGGAACCTGCATCTAGCTCGT 755	
QY	113	AGCACCGGGTGTCTGTGCGCGCTCCCGCTGTTCCACAGCCCTTCGCACATCGCCTGCGTCC 172	
DB	756	CGACGTGCTGCTGTGGGTGTGCGGTGTTCCACATCTACTCTCACTCACTCGTGTGTCG 815	
QY	173	TGTCGCGCCACCGTGGTTCGGGCGACAGCTTCGGATTCGCGAGCGGCGCTCCGCGACGACG 232	
DB	816	TCGCGGCTCTCCGCGCGGGTGCAGATCGTATCATGCGCAAGTTCGACCCACCGCGCGC 875	
QY	233	TGATCGCGCTGATCAGGCGGAGAGCTGCACCTTCCTGGCGGGGTGCGACCACTTACC 292	
DB	876	TGTTGGACCTTGGTTCGACGACGCGCTCACCTGTGGGCGCCATTCGTGCGCGCCATCGTGG 935	
QY	293	ACCACTTGGTTCGGGCGCGCGCGCGCGGCTTTCTCGCGCCGAGCTGCGGATCGCGCC 352	
DB	936	TGAGATTCGCAAGAGCGGCGGGTGACCGCGCGGACCTGGCTTCATCCGGCTGTGCA 995	
QY	353	TGGCGCGGGCGCGGTCTCTCGGCGCGGGCTGTGCGAGCGAGTTC---GAAGACCTTCG 409	
DB	996	TGTGCGGGCGCGGCCCATGTCGCAAGAGCTGCAGAGCGCGTTCATGCGCAAGATCCCCA 1055	
QY	410	GGGTCCCGCTGATCGACGCTACGGCAGACCGAGACCTGCGGGCGATCACCATTGAACC 469	

Db	1056	ACGCGTGTCTGGCCAGGATATGGATGACCGAGCGCGCCTGTGCTGGCGATGTGCC 1115
QY	470	CGCCGGACGGC 480
Db	1116	TGGCCTTCGCC 1126
RESULT 14		
ABK13573		
ID	ABK13573	standard; cDNA; 2038 BP.
XX	AC	ABK13573;
XX	DT	23-APR-2002 (first entry)
XX	DE	Ryegrass 4-coumarate Co-A-ligase 3 (Lp4CL3) cDNA.
XX	XX	Perennial ryegrass; ss; lignin; 4 coumarate CoA-ligase; QTL;
XX	XX	lignin biosynthesis; enzyme; cinnamoyl-CoA reductase; CCR; Lp4CL3;
XX	XX	cinnamyl alcohol dehydrogenase; CAD; molecular genetic marker; gene;
XX	XX	qualitative trait loci; tagging; QTL mapping; DNA fingerprinting;
XX	XX	marker assisted selection; forage improvement; turf grass improvement;
XX	XX	dry matter digestibility; herbage quality; palatability; regrowth;
XX	XX	cold tolerance; drought tolerance; tiller survival; plant persistence.
XX	XX	Lolium perenne.
XX	XX	Key
XX	5'UTR	1. .112
XX	FT	/tag= a
XX	FT	113. .1786
XX	FT	/tag= b
XX	FT	/product= "Lp4CL3 protein"
XX	3'UTR	1787. .2019
XX	FT	/tag= c
XX	FT	2020. .2038
XX	FT	/tag= d
XX	PN	WO200195702-A1.
XX	XX	20-DEC-2001.
XX	XX	14-JUN-2001; 2001WO-AU000699.
XX	XX	14-JUN-2000; 2000AU-00008154.
XX	XX	(VICT-) STATE VICTORIA DEPT NATURAL RES & ENVIRO.
XX	XX	(UYAD-) UNIV ADELAIDE.
XX	XX	(ITMA-) INT MAIZE & WHEAT IMPROVEMENT CENT.
XX	XX	(SAUS-) STATE SOUTH AUSTRALIA SOUTH AUSTRALIAN R.
XX	XX	(UYSC-) UNIV SOUTHERN CROSS.
XX	XX	(DAIR-) DAIRY RES & DEV CORP.
XX	XX	Spangenberg GC, Lidgett AJ, Heath RL, McInnes RL, Lynch DP;
XX	XX	WPI; 2002-097993/13.
XX	XX	P-PSDB; AAU75085.
XX	XX	Novel nucleic acid encoding enzymes involved in lignin biosynthetic
XX	XX	pathway from ryegrass or fescue species useful for modifying lignin
XX	XX	biosynthesis in plants and as a molecular genetic marker.
XX	XX	Claim 3; Fig 4; 148pp; English.
XX	XX	This invention represents purified or isolated nucleic acid and protein
XX	XX	sequences of enzymes involved in lignin biosynthesis. The enzymes of the
XX	XX	invention are 4 coumarate CoA-ligase (4CL), cinnamoyl-CoA reductase (CCR)
XX	XX	and cinnamyl alcohol dehydrogenase (CAD) from a ryegrass (Lolium sp.) or
XX	XX	fescue (Festuca sp.). The invention also comprises an isolated regulatory
XX	XX	element from the nucleic acid sequences and a plant cell or seed
XX	XX	transformed with the nucleic acid. An isolated regulatory element from
XX	XX	these nucleotide molecules is useful for expressing an exogenous gene in

CC plant cells. The nucleotide sequences of the invention and vectors  
CC containing these sequences are useful for modifying lignin biosynthesis  
CC in a plant and are useful as a molecular genetic marker for qualitative  
CC trait loci (QTL) tagging, QTL mapping, DNA fingerprinting and in marker  
CC assisted selection, in forage and turf grass improvement, e.g. tagging  
CC QTLs for dry matter digestibility, herbage quality, palatability,  
CC regrowth after cutting and grazing, cold tolerance, drought tolerance,  
CC tiller survival and plant persistence. The present sequence represents  
CC the perennial ryegrass 4-coumarate Co-A-ligase 3 (Lp4CL3) cDNA of the  
CC invention  
XX  
SQ Sequence 2038 BP; 416 A; 627 C; 627 G; 368 T; 0 U; 0 Other;

Query Match 16.4%; Score 82.2; DB 6; Length 2038;  
Best Local Similarity 51.1%; Pred. No. 1.4e-05;  
Matches 251; Conservative 0; Mismatches 228; Indels 12; Gaps 2;  
QY 2 TGTTCACAGTCCGGGACGACGCGGGGCGCCCAAGGGCGTGTGTCACGAGCGCAACT 61  
DB 693 TGCCTACTCTCCGGACACGCGGGCTCCCAAGGGCGTCATGCTCACCCACCGCAGCC 752  
QY 62 GCCTCTGGTGGTGGCTTC-----CTGCTAGTGGCTTCCCGGGTGTGCGACC 112  
DB 753 TCGTCACACGCGTCCCGCAGCTGGTCGACGCGGTGCAACCCCTTAACGTGTCTCAACAAGG 812  
QY 113 AGGACCGGGTGTCTGGCGGCTCCCGCTGTTCCACAGCTTTCACATCGCTCGCTGCC 172  
DB 813 ACGACGGCTGTGTGTGCTGTGGCTTTCACATCTACTCTGCTGCACAGGTGTGTC 872  
QY 173 TGTCCGCAACGTTGGTCCGGGCCAGCGTCCGGATCGCGAGCTTCGCCCGCAGCAGC 232  
DB 873 TGGCGGGGCTCCGGTGTGGCGCGCCATCGTCATCGGCAAGTTCGACGTCCGGCGGC 932  
QY 233 TGATCGGCTCATCAGGCGAGAGCTCGACCTTCTTGGCGGGTGGCGACCACTTACC 292  
DB 933 TGGTGGACCTCGTCCGGCGCACCGCATACCATCGCGCCATTCTGTCCCGCCCTCGTCG 992  
QY 293 ACCACCTGGTCCGGCGCGCCCGGCGAGCGCGTTCCTCCGCGCCGAGCTCGCGATCGGCC 352  
DB 993 TGGAGATCGCCAAAGAGACCGCGTCCGGCGCGAGGACCTCGCATCATCGCATGTGC 1052  
QY 353 TGGCCGGGGCGCGGTCTCTCGCGCGCGGGCTCGAAGCGAGTTC---GAAGAGACCTTCG 409  
DB 1053 TCTCCGGCGCGCGCCATGCGGCAAGGACCTCCAGAGCGCTTCATGSCCAAGATCCCA 1112  
QY 410 GGGTCCCGCTCATCGACGCTTACGCGACGCGAGCTCGGGGCGCATCACCATGACCC 469  
DB 1113 ACGCGGTGCTCGACAGGGGTACGGGATGACCGAGGCTGGGCGCGGTGTGCGCCATGTGCC 1172  
QY 470 CGCGCGAGCGGC 480  
DB 1173 TGGCGTTGCC 1183

RESULT 15  
AD68539 standard; cDNA; 2073 BP.  
XX  
AC AD68539;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE S. arundinaceus lignin/tannin biosynthesis protein cDNA SEQ ID NO:31.  
XX  
KW Lolium perenne; Festuca arundinacea; lignin; fructan; tannin;  
XX biosynthetic pathway; plant; gene; ss.  
XX  
OS Schedonorus arundinaceus.  
XX  
PN WO2003040306-A2.  
XX  
PD 15-MAY-2003.  
XX

PF 07-NOV-2002; 2002WO-NZ000239.  
XX  
PR 07-NOV-2001; 2001US-0337703P.  
XX  
PA (GENE-) GENESIS RES & DEV CORP LTD.  
PA (WRIG-) WRIGHTSON SEEDS LTD.  
XX  
PI Demmer J, Forster RL, Gibson JB, Shenk MA, Norriess MG, Glenn M;  
PI Saulebury KM, Hall C;  
XX  
DR NPI; 2003-441544/41.  
DR P-PSDB; ADC68383.  
XX  
PT New polynucleotide encoding polypeptides from Lolium perenne or Festuca  
PT arundinacea, useful for modulating the biosynthesis of lignin, fructan or  
PT tannin in a plant.  
XX  
PS Claim 1; SEQ ID NO 31; 240pp; English.  
XX  
CC The present invention describes isolated polynucleotides (I) encoding  
CC proteins (II) from Lolium perenne and Festuca arundinacea which are  
CC active in lignin, fructan and tannin biosynthetic pathways. Also  
CC described: (1) an isolated oligonucleotide probe or primer comprising at  
CC least 10 contiguous residues complementary to 10 contiguous residues of  
CC (1); (2) a kit comprising the oligonucleotide probe or primer; (3) a  
CC genetic construct comprising (I); (4) a transgenic plant cell comprising  
CC the genetic construct of (3); (5) a plant or its seed, fruit or progeny  
CC comprising the transgenic plant cell of (4); (6) modulating one or more  
CC of the lignin, fructan or tannin compositions of a plant; (7) producing a  
CC plant having one or more of the lignin, fructan or tannin compositions;  
CC and (8) modifying the activity of (II) involved in a lignin, fructan or  
CC tannin biosynthetic pathway in a plant. (I) can be used for modulating  
CC the biosynthesis of lignin, fructan or tannin in a plant. The present  
CC sequence is used in the exemplification of the present invention.  
XX  
SQ Sequence 2073 BP; 430 A; 646 C; 627 G; 370 T; 0 U; 0 Other;

Query Match 16.4%; Score 82.2; DB 9; Length 2073;  
Best Local Similarity 51.1%; Pred. No. 1.4e-05;  
Matches 251; Conservative 0; Mismatches 228; Indels 12; Gaps 2;  
QY 2 TGTTCACAGTCCGGGACGACGCGGGCGGCCCAAGGGCGTGTGTCACGAGCGCAACT 61  
DB 717 TGCCTACTCTCCGGACACGCGGGTCCCAAGGGCGTCATGCTCACCCACCGCAGCC 776  
QY 62 GCCTCTGTCTCGTCG-----TTCCTCTAGCTGCGGTTCCTCCGGGTGTGCGACC 112  
DB 777 TCGTCACAGCGTCCCGCAGCTGTTGACGGGTGACACCCGAAAGTGTGCTTCACACAGG 836  
QY 113 AGGACCGGGTGTCTCTGGCGCGTCCCGCTGTTCCACAGCTTTCGACATCGCTGCGTCC 172  
DB 837 ACGACGCGTGTGTGCTGCTGCGGTCTTCCACATCTACTCTGTCACACGCTGCTGC 896  
QY 173 TGTCCGCCACCGTGTCCGGGCGAGCTCCGGATTCGCGAGCTCCCGCCGACGAGC 232  
DB 897 TGGCGGGGCTCCGGTGTGGGCGCCCATCGTCATCGCAAGTTCGACGTCCGCGCGC 956  
QY 233 TGATCGGCTGATCGAGCGGAGAGCTCGACTTCTTCTGGCGGGTTCGCGACCACTACC 292  
DB 957 TGGTGGACCTCGTCCGGCGCACCGCATCACATGCGGCCATTGTGCCCCCGATCGTGG 1016  
QY 293 ACCACCTGTGTCCGGCGCGCCCGCGCAGCGCGGTTCCTCCGCGCCGAGCTTCGCGATCGGCC 352  
DB 1017 TGGAGATCGCCAAAGAGGACGCGCTCACCGCGACGCTCACATCCATCCGATGTGTC 1076  
QY 353 TGGCGGGGCGCGGTCTCGCGCGGGCTGCGAAGCGAGTTC---GAAGAGACCTTCG 409  
DB 1077 TCTCCGGCGCGCGCCCATGGGCAAGGACCTCCAGGACGCGTTCATGSCCAAGATCCCA 1136  
QY 410 GGGTCCCGCTCATCGACGCTTACGCGAGACGACGAGAGCTCGGGGCGCATCACCATGAC 469  
DB 1137 ACGCGGTCTCGGACAGGGTTACGGGATGATGAGGCGCGGTCCGGTGTGCGCATGTGCC 1196

Qy 470 CGCCGACGGC 480  
| | | | |  
Db 1197 TGGCGTTGCC 1207

Search completed: June 16, 2004, 23:46:37  
Job time : 290 secs

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	184.2	36.8	77536	4	US-09-410-551B-1	Sequence 1, Appli
C 2	84.2	16.8	1584	4	US-09-252-991A-14522	Sequence 14522, A
C 3	84.2	16.8	1806	4	US-09-252-991A-14654	Sequence 14654, A
C 4	84.2	16.8	1365	4	US-09-252-991A-15096	Sequence 15096, A
C 5	80.8	16.1	1749	4	US-09-489-039A-805	Sequence 805, App
C 6	78.4	15.6	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C 7	78.4	15.6	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C 8	76.6	15.3	30001	2	US-08-125-468-1	Sequence 1, Appl
C 9	76.6	15.3	30001	2	US-08-474-933-1	Sequence 1, Appl
C 10	69.2	13.8	1386	4	US-09-252-991A-3821	Sequence 3821, Ap
C 11	69.2	13.8	1282	4	US-09-252-991A-3777	Sequence 3777, Ap
C 12	69.2	13.8	2181	4	US-09-252-991A-3690	Sequence 3690, Ap
C 13	66	13.2	2124	4	US-09-266-965-44	Sequence 44, Appl
C 14	66	13.2	53500	4	US-09-266-965-76	Sequence 76, Appl
C 15	61.6	12.3	2043	4	US-09-615-192A-866	Sequence 266, App
C 16	61.6	12.3	5103	4	US-09-252-991A-5192	Sequence 5192, Ap
C 17	61.6	12.3	6876	4	US-09-252-991A-5283	Sequence 5283, Ap
C 18	61	12.2	8051	2	US-08-576-626A-2	Sequence 2, Appl
C 19	60.2	12.0	957	4	US-09-252-991A-9229	Sequence 9229, Ap
C 20	60.2	12.0	1182	4	US-09-252-991A-9062	Sequence 9062, Ap
C 21	60.2	12.0	1209	4	US-09-252-991A-8718	Sequence 8718, Ap
C 22	59.4	11.9	543	4	US-09-615-192A-197	Sequence 197, Appl
C 23	59.2	11.8	357	4	US-09-489-039A-706	Sequence 706, App
C 24	58.6	11.7	1152	4	US-09-252-991A-7729	Sequence 7729, Ap
C 25	58.6	11.7	1215	4	US-09-252-991A-7837	Sequence 7837, Ap
C 26	58.8	11.7	1581	4	US-09-252-991A-7755	Sequence 7755, Ap
C 27	58.2	11.6	1320	4	US-09-252-991A-3999	Sequence 3999, Ap





Db 1213 C 1213

RESULT 4  
US-09-252-991A-15096/c  
; Sequence 15096, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: 107196.136  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15096  
; LENGTH: 1965  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15096

Query Match 16.8%; Score 84.2; DB 4; Length 1965;  
Best Local Similarity 49.9%; Pred. No. 2.1e-08;  
Matches 240; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

Qy 7 TACACGTCGGGCAAGACCGGCGGCCCAAGGGGCTGTCGTCACGACGCGCAACTGCTC 66  
Db 1305 TACACCTCGGTATACACCGGCTTCCCAAGGGCGCACCTCAGCCACTACAACATCCTC 1246

Qy 67 TGGTCCGTGCTCTCTGCTAGTCGCTGCGGTTCCCGGGTGTGGACCGAGACGGGTGCTC 126  
Db 1245 AACAAACGGCTACATGCTGCGGGAAGCCTCGGACTCGGCGCGGAAGACCGCTGCTGATC 1186

Qy 127 TGGCGGCTCCCGCTGTTCCACAGCCTTTCCGACATCGCTGCTCTGTCGCGCACCGGTG 186  
Db 1185 CCGGTGCGCTGATACATGCTTGGCATGTTGATGGCAACCTCGGTGCTGATCACCAC 1126

Qy 187 GTCGGGGCCAGCTCCGATCCGACGCGGAGCTCCGCGGACGATGATGGGTGATC 246  
Db 1125 GGTTCGACCATGATACCTCCGGGCGGAGCTTCGACGCGGAGGACCTGCTGCGCGTG 1066

Qy 247 GAGCGGAGAGCTGACCTTCTGCGCGGCTGCGGACCACTACCACTGCTGCGG 306  
Db 1065 GCCGAGGAGCGCGCACCGCTGTACCGGGGTACCGGAGTACCGACATGTTTCATCGCGGAGCTGGAT 1006

Qy 307 GCGCGCGCGGAGCGGGTCTTCTCCGCGCGGAGCCTCGCGATCGG---CTGCGCGGGGCG 363  
Db 1005 CATCGCGCGCGCGGATTCGACCTTTCAGCCTTCCGCGGACCGGATCATGGCGGCGCC 946

Qy 364 CGGTCTCGGCGCGGCTCGAGAGCGAGTTGGAAGAGACTTCGGGGTCCCGTGTATC 423  
Db 945 ACCTGCCGATCGAGGTGATGCGGGGGTTCATTGGCGCATGCATGACATGCGCGAGGTGCG 886

Qy 424 GAGCGCTAGGAGAGCAGCGGAGCTGCGGGGCGATCACCATGAACCGCGGAGCGCGCC 483  
Db 885 ATCGCGCTAGGAGATGACCGGAGACCACTCGCGGTGCTGCTGACAGCGGTCCCGACGATGGC 826

Qy 484 C 484  
Db 825 C 825

RESULT 5  
US-09-489-039A-805  
; Sequence 805, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 805  
; LENGTH: 1749  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-805

Query Match 16.1%; Score 80.8; DB 4; Length 1749;  
Best Local Similarity 49.0%; Pred. No. 9.5e-08;  
Matches 245; Conservative 0; Mismatches 252; Indels 3; Gaps 1;

Qy 2 TGTTCACAGTCGGGCAAGACCGGCGGCCCAAGGGGCTGTCGTCACGACGCGCAACT 61  
Db 665 TCTGTACACCTCGGAAACACACCGCGCGCGGAAAGGGGTATTGAACACCACTCGTTGCG 724

Qy 62 GCCTCTGGTCGCTGCTTCTGCTAGCTGCG---TTCCCCGGGTGTCGACACGAGACC 118  
Db 725 TGGTCTCTGACACGCTTGAAGCGCAACACGCGCGGACGCGCGGCATATCGGCGAGGACA 784

Qy 119 GGGTGTCTTGCGCGCTCCCGCTGTTCCACAGCCTTTCGCACATCGCCTGCTGCTGTCG 178  
Db 785 GCCTGTTCCCGGTGTCGCGATGTTCCATGTCACGCGTGGGGGACGCCGTTTATCGCG 844

Qy 179 CCACGCTGTCGGGGCCAGCTCGCGATCGCGGACGCGGAGCTCCGCGGACGACGATGC 238  
Db 845 CCATGCTCGGCGCGGCTGCTGCTGCGCGGCGGACCTCGATGCGCGACGACCTGCTGC 904

Qy 239 GGCTGATCGAGCGGAGCTCGACCTTCTTCCGCGCGGCTGCGGACCTACCACTACCACT 298  
Db 905 AGCTGCTGCGCGGAGAGGTGACGCTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGG 964

Qy 299 TGGTGGCGCGCGCGCGCGCGCTTCTCCGCGCGGAGCTGCGGATGCGGCTGCGGCTG 358  
Db 965 TGCTGGCGCGCATGCGCGCGGACAGGTTCGCGCTGCTGATTTAAAGTGGCTGGTGG 1024

Qy 359 GGGGCGCGGCTCTCGCGCGCGGCTGCGAAGCGAGTTCGAAGAGACCTTCGCGGGTCCCG 418  
Db 1025 GCGGCTCGCGCTGCGCGCGTGCATGGCGGAGGCTTTCAGCGTGCATTAACGCAATTGCAC 1084

Qy 419 TGATCGAGCTACGCGGACGCGGACCTCGGCGGCGATCACCATGAACCGCGCGGAGC 478  
Db 1085 TGACCCAGCTGGGGAATGACCGGAAACATCGCCCATCGGCACTACACACCGTTGA 1144

Qy 479 GCGCGCGCGCTCGAGGGCTCG 498  
Db 1145 GCACACATGACGCGCTCCCG 1164

RESULT 6  
US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765



```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

Query Match          15.3%; Score 76.6; DB 1; Length 30001;
Best Local Similarity 52.7%; Pred. No. 6.4e-07;
Matches 166; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 176 CCGCCACCGTGTGCGGGCCAGCGTCGGAGTCGCGGACGCGGAGCTCCGCCGACGACGTGA 235
Db 9748 CCGCGCTCGCGGTGCGGCGGACGACGCTGTCTGCCCGGAGAGGACCGCGCGCGCTGG 9807

QY 236 TGGGCTGATCGAGCGGAGAGCTCGACCTTCTTGGCGCGGCTGCCGACCTTACACACC 295
Db 9808 TGGGACTGCGCGGTACCTTGGCGGCGCTCGACCTGTACAGCTGCGCGGTGTGTCCC 9867

QY 296 ACCTGGTGGCGGCGCGCGCGGCTTCTCGCGCGGAGCTTGGGAGTCCGCGCTGG 355
Db 9868 GGCTCGCGGCGGAGCGAGCGGTGCGCGCTTCCGCTTCCGATCCCTCGCTACGCGCGGTGTGT 9927

QY 356 CCGGGGGCGCGGTCTCTCGCGCGCGGCTGCGAAGCGAGTTGCGAAGAGACCTTTCGGGGTCC 415
Db 9928 CCGGTGGATCGCGCTGCGCGCGCGGACGACGCGCGCTGCGAGGACCTTTCGGGGTGC 9987

QY 416 CGCTGATCGAGCTTACGCGAGCAGCGGAGACCTTGGGGGCGATCACCATGACCGCGCG 475
Db 9988 CGGTGCTCCAGGCTTACGCGCTGCGCGGAGACGTCGCCGTCGAGCAGCTTCGACCTGCCCCG 10047

QY 476 ACGGCGCGCGCGTGC 490
Db 10048 AGGGGCCACCCCTCG 10062

RESULT 9
US-08-474-933-1
; Sequence 1, Application US/08474933
; Patent No. 5866410
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy
; APPLICANT: Fantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmid
; TITLE OF INVENTION: useful therein
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,933
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,468
; FILING DATE: 22-SEP-1993
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match          15.3%; Score 76.6; DB 2; Length 30001;
Best Local Similarity 52.7%; Pred. No. 6.4e-07;
Matches 166; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 176 CCGCCACCGTGTGCGGGCCAGCGTCGGAGTCGCGGACGCGGAGCTCCGCCGACGACGTGA 235
Db 9748 CCGCGCTCGCGGTGCGGCGGACGACGCTGTCTGCCCGGAGAGGACCGCGCGCGCTGG 9807

QY 236 TGGGCTGATCGAGCGGAGAGCTCGACCTTCTTGGCGCGGCTGCCGACCTTACACACC 295
Db 9808 TGGGACTGCGCGGTACCTTGGCGGCGCTCGACCTGTACAGCTGCGCGGTGTGTCCC 9867

QY 296 ACCTGGTGGCGGCGCGCGGCTTCTCGCGCGGAGCTTGGGAGTCCGCGCTGG 355
Db 9868 GGCTCGCGGCGGAGCGAGCGGTGCGCGCTTCCGCTTCCGATCCCTCGCTACGCGCGGTGTGT 9927

QY 356 CCGGGGGCGCGGTCTCTCGCGCGCGGCTGCGAAGCGAGTTGCGAAGAGACCTTTCGGGGTCC 415
Db 9928 CGGTGCTCCAGGCTTACGCGCTGCGCGGAGACGTCGCCGTCGAGCAGCTTCGACCTGCCCCG 10047

QY 416 CGCTGATCGAGCTTACGCGAGCAGCGGAGACCTTGGGGGCGATCACCATGACCGCGCG 475
Db 9988 CGGTGCTCCAGGCTTACGCGCTGCGCGGAGACGTCGCCGTCGAGCAGCTTCGACCTGCCCCG 10047

QY 476 ACGGCGCGCGCGTGC 490
Db 10048 AGGGGCCACCCCTCG 10062

RESULT 10
US-09-252-991A-3821
; Sequence 3821, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3821
; LENGTH: 1986
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3821

Query Match          13.8%; Score 69.2; DB 4; Length 1986;
Best Local Similarity 55.8%; Pred. No. 1.7e-05;
Matches 153; Conservative 0; Mismatches 118; Indels 3; Gaps 1;

QY 225 CGACGACGTGATGCGGCTGATCGAGCGGAGAGCTTCGACCTTCCTGGCGCGCGTCCGAC 284
```

Db 987 CCAGGACTTCTGGAAGCTGGTGGAAAGCTACCGGGTCCACAGCTTCAGTGGGTGCGGAC 1046  
 QY 285 GACCTACACACCTGCTGGTGGGGGCGCCGCGAGCGGGTTCCTCGCGCCGAGCTGCG 344  
 Db 1047 CATCTACGCGGGCTGCTGCTGAGTCTCCGAGCGATGCGCG---CGACCTGTCCAGCCTGG 1103  
 QY 345 GATCGGCTGGCGGGGCGCGCTCTCGCGCGCGGGCTGCGAAGCGAGTTCGAAGAGAC 404  
 Db 1104 TTTGCGCCCTGTGGGCGCGCGCGCGATGCGGTGGAGCTGATCCGCGAGTTCAGGCGCG 1163  
 QY 405 CTTGGGGTCCCGCTGATCGACGCTTACGCGACCGAGACCTCGCGGGCGATCACCAT 464  
 Db 1164 TACCGGGCTCAAGGTGATCGAGGGCTACGCGCTCACCGAAGGTACCTGCGGACCAAGTTG 1223  
 QY 465 GAACCCGCGGAGCGCGCGCGCTCGAGGGCTCG 498  
 Db 1224 CAACCCGCGTGGCGCGAGCGCGCGCGGGTTGCG 1257

RESULT 11  
 US-09-252-991A-3777  
 ; Sequence 3777, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 3777  
 ; LENGTH: 2082  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-3777

Query Match 13.8%; Score 69.2; DB 4; Length 2082;  
 Best Local Similarity 55.8%; Pred. No. 1.7e-05;  
 Matches 153; Conservative 0; Mismatches 118; Indels 3; Gaps 1;  
 QY 225 CGAGGACTGATGCGGCTGATCGAGGCGGAGAGCTCGACCTTCTCGCGCGGCTGCGGAC 284  
 Db 1156 CCAGGACTTCTGGAAGCTGGTGGAAAGCTACCGGGTCCACAGCTTCAGTGGCGTGGCGAC 1215  
 QY 285 CACCTACACACCTGCTGGTGGGGGCGCGCGCGAGCTGCTTCGCGCGGAGCCTGCG 344  
 Db 1216 CATCTACGCGGGTGTCTGAGGTCCCGAGGTCCCGAGTGGCG---CGACCTGTCCAGCTGCG 1272  
 QY 345 GATCGGCTGGCGGGGCGCGGTCTCGCGCGCGCGGTGCGAAGCGAGTTCGAAGAGAC 404  
 Db 1273 TTTGCGCCCTGTGGCGCGCGCGCGATGCGCGGTGAGCTGATCGCGCACTTCGAGGCGCG 1332  
 QY 405 CTTGGGGTCCCGCTGATCGACGCTTACGCGACCGAGACCTCGCGGGCGATCACCAT 464  
 Db 1333 TACCGGGCTCAAGGTGATCGAGGGCTACGCGCTCACCGAAGGTACCTGCGGCGCAAGTTG 1392  
 QY 465 GAACCCGCGGAGCGCGCGCGCTCGAGGGCTCG 498  
 Db 1393 CAACCCGCGTGGCGCGAGCGCGCGCGGGTTGCG 1426

RESULT 12  
 US-09-252-991A-3690/c  
 ; Sequence 3690, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 3690  
 ; LENGTH: 2181  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-3690  
 Query Match 13.8%; Score 69.2; DB 4; Length 2181;  
 Best Local Similarity 55.8%; Pred. No. 1.7e-05;  
 Matches 153; Conservative 0; Mismatches 118; Indels 3; Gaps 1;  
 QY 225 CGAGGACTGATGCGGCTGATCGAGGCGGAGAGCTCGACCTTCTCGCGCGGCTGCGGAC 284  
 Db 1015 CCAGGACTTCTGGAAGCTGGTGGAAAGCTACCGGGTCCACAGCTTCAGTGGCGTGGCGAC 956  
 QY 285 CACCTACACACCTGCTGGTGGGGGCGCGCGCGAGCGGTTCCTCGCGCGGAGCCTGCG 344  
 Db 955 CATCTACGCGGGCTGCTGCTGAGGTCCCGAGCGATGCGCG---CGACCTGTCCAGCCTGG 899  
 QY 345 GATCGGCTGGCGGGGCGCGGTCTCGCGCGCGCGGTGCGAAGCGAGTTCGAAGAGAC 404  
 Db 898 TTTGCGCCCTGTGCGCGCGCGCGCGATGCGGTGGAGCTGATCCGCGAGTTCAGGCGCG 839  
 QY 405 CTTGGGGTCCCGCTGATCGACGCTTACGCGACCGAGACCTCGCGGGCGATCACCAT 464  
 Db 838 TACCGGGCTCAAGGTGATCGAGGGCTACGCGCTCACCGAAGGTACCTGCGGCGCAAGTTG 779  
 QY 465 GAACCCGCGGAGCGCGCGCGCTCGAGGGCTCG 498  
 Db 778 CAACCCGCGTGGCGCGAGCGCGCGGGTTGCG 745

RESULT 13  
 US-09-266-965-44  
 ; Sequence 44, Application US/09266965  
 ; Patent No. 6495348  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D  
 ; APPLICANT: Mao, Y  
 ; APPLICANT: Varoglu, M  
 ; APPLICANT: He, M  
 ; APPLICANT: Sheldon, P  
 ; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
 ; FILE REFERENCE: 600.456US1  
 ; CURRENT APPLICATION NUMBER: US/09/266,965  
 ; CURRENT FILING DATE: 1999-03-12  
 ; EARLIER APPLICATION NUMBER: US 08/624,447  
 ; EARLIER FILING DATE: 1996-08-19  
 ; EARLIER APPLICATION NUMBER: PCT/US94/11279  
 ; EARLIER FILING DATE: 1994-10-06  
 ; EARLIER APPLICATION NUMBER: US 08/133,963  
 ; EARLIER FILING DATE: 1993-10-07  
 ; NUMBER OF SEQ ID NOS: 145  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 44  
 ; LENGTH: 2124  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces lavendulae  
 US-09-266-965-44

Query Match 13.2%; Score 66; DB 4; Length 2124;  
 Best Local Similarity 49.6%; Pred. No. 7.2e-05;  
 Matches 233; Conservative 0; Mismatches 225; Indels 12; Gaps 2;  
 QY 2 TGTCTACGTCGGGCGGACCGCGGCGCCCAAGGGCGTGTGTCCACGAGCGCAACT 61

Db 233 TGCTCCGACCTCGGGTTCAGCCGTCGAGCAAGCTCGTCCGCGCAGCGAGAGACC 292  
QY 62 GCCTCTGGTCCGTCTCTCTGCTA CTTGCGGTTCCTCCGGTTCGAGACCGGG 121  
Db 293 TGCTCCGAGAGCGCGCGCTTACCGCAGCGGGT---CGGGCTGACCGGAGAGACACC 349  
QY 122 TGCTCTGGCGCTCCCGCTCTTCCAC-----AGCCTTTGACATGCGCTGCGTCC 172  
Db 350 TGCTCTGGCGCTCCCGCTCTTCCACCGGTCAGCGCTGCGGTCTGCTGCGGAGTGC 409  
QY 173 TGTCCGACACCGTCTGGGGCCAGCGCTTCGAGTCGCGGACGCGAGCTTCGCGCAGCAG 232  
Db 410 TGACGGGTGCGCGCTGCGCCCGTACCGCGCAGCGCCCTCGCGCGGATCGCGCGAGC 469  
QY 233 TGATGCGGCTGATCGAGCGGAGAGCTTGACCTTCTTGGCGCGGTGCGGACCACTTACC 292  
Db 470 TGTCCGCTGCTGCGACCGTGTGCGCTCTGCTGCGCTGCTGCGCGCGTGTGCGCGACC 529  
QY 293 ACCACCTGCTGCGGGCGCGCCCGCAGCGCGTTCCTCGCGCGCGAGCTGCGGATCGGCC 352  
Db 530 GCGCGCTGCGGGGAGACGCGCGCGCGCGCGCGCGCGCTTCCACCGAGCGTTCGCGGA 589  
QY 353 TGGCGCGGGCGCGCTCTCGCGCGCGCGCTGCGAAGCGAGTTCGAAAGAGACTTCGCGG 412  
Db 590 TGGTGGGTGCGGGCGCGCTGCGAGCAGCAGCTGCGACCGCGCTTACCGAGCGTTCGCGGA 649  
QY 413 TCCGCTGATCGACCGCTTACCGCAGCAGCAGCAGCTGCGGGCGGATCACC 462  
Db 650 CCGGTCTCGCGCGCAACTACGCTTCCACGAGAGCGGGCGCGTGTCTGCC 699

RESULT 14  
US-09-266-965-76/c  
; Sequence 76, Application US/09266965  
; Patent No. 6495348  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456U1  
; CURRENT APPLICATION NUMBER: US/09/266,965  
; EARLIER FILING DATE: 1999-03-12  
; EARLIER APPLICATION NUMBER: US 08/624,447  
; EARLIER FILING DATE: 1996-08-19  
; EARLIER APPLICATION NUMBER: PCT/US94/11279  
; EARLIER FILING DATE: 1994-10-06  
; EARLIER APPLICATION NUMBER: US 08/133,963  
; EARLIER FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 76  
; LENGTH: 53500  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-266-965-76

Query Match 13.2%; Score 66; DB 4; Length 53500;  
Best Local Similarity 49.6%; Pred. No. 7.4e-05;  
Matches 233; Conservative 0; Mismatches 225; Indels 12; Gaps 2;  
QY 2 TGTTCACAGTCGCGGACAGACCGGCGCGCCCAAGCGGCTGTGTTCACGCGCGCACT 61  
Db 16979 TGTCCCGACCTCGGGTTCGACCGGTGCGAGCAAGCTCGTTCGCGCGCAGCGAGAGCC 16920  
QY 62 GCCTCTGGTCCGTCTCTCTGCTA CTTGCGGTTCCTCCGGTTCGAGACCGGG 121  
Db 16919 TGCTCGGAGGCGCGCGCTTACCGGACGGGT---CGGGCTGACGGGAGAGACACC 16863  
QY 122 TGCTCTGGCGCTCCCGCTCTTCCAC-----AGCCTTTGACATGCGCTGCGTCC 172

Db 16862 TGCTCTGCGGTGCGCGCTGTCCACCGCTACGCGCTGGGCTGGCTGTTCGGCGGACTGC 16803  
QY 173 TGTCCGCGACCGTGTTCGGGGCCAGCGCTCCGAGTCGCGCGGAGCTTCGCGCGGACGACG 232  
Db 16802 TGACGGGTGCGCGCTGCGCCCGTACCGCGCAGCGCGCTTCGGCGCATCGCGCGGAGC 16743  
QY 233 TGATGCGGCTGATCGAGCGGAGAGCTTCGACCTTCTTGGCGCGGTGCGGACCACTTACC 292  
Db 16742 TGTCCGCTGCTGCGACCGTGTGCGCTCTGCTGCGCAGTGTGCGCGCGTGTCTGCGGACCC 16683  
QY 293 ACACCTGCTGCGGGCGCGCCCGCAGCGCGGTTCCTCGCGCGGAGCTTCGCGGATCGGCC 352  
Db 16682 GCGCGCTGCGGGGAGCAGCGCGCGCGCGCGCGCTTCCCGGTCTCCCGGTTCGCGCTGGCCA 16623  
QY 353 TGGCGCGGGCGCGGTCTCTCGCGCGCGCGCTGCGAAGCGAGTTCGAAAGAGACTTCGCGG 412  
Db 16622 TGGTGGGTGCGGGCGCGGTGAGCAGCAGCTGCGACCGCGGTTCACCGAGGCGTTCGCGA 16563  
QY 413 TCCGCTGATCGACCGCTTACCGCAGCAGCAGCAGCTTCGCGGGCGGATCACC 462  
Db 16562 CCGGTCTCGCGCGCAACTACGCTTCCACGAGAGCGGGCGCGTGTCTGCC 16513

RESULT 15  
US-09-615-192A-266  
; Sequence 266, Application US/09615192A  
; Patent No. 6410718  
; GENERAL INFORMATION:  
; APPLICANT: Bioskera, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; MODIFICATION OF PLANT LIGNIN CONTENT  
; FILE REFERENCE: 11000.1003c4U  
; CURRENT APPLICATION NUMBER: US/09/615,192A  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 08/975,316  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: US 08/713,000  
; PRIOR FILING DATE: 1996-09-11  
; PRIOR APPLICATION NUMBER: US 09/169,789  
; PRIOR FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 405  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 266  
; LENGTH: 2043  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-09-615-192A-266

Query Match 12.3%; Score 61.6; DB 4; Length 2043;  
Best Local Similarity 48.8%; Pred. No. 0.00052;  
Matches 234; Conservative 0; Mismatches 234; Indels 12; Gaps 2;  
QY 6 CTACACGTGCGGACAGCAGCGCGCGCCCAAGCGGTGTGTTCACGCGCG----- 56  
Db 649 CTATTGCTGCGGACAGCAGCGGGCTTCCCAAGGAGTGTATGCTTACGACAGGGGTCAAGT 708  
QY 57 GAACCTGCTCTGCTCGGTCTCTCTGCTACGTCGCGGTTCGCCGGGTTCGCGACGAGGA 116  
Db 709 GACCAAGCTGCGCGCAGCAGAGTTCGAGGAGAACCCCAACTGTACTTCCACAGAGAGA 768  
QY 117 CCGGGTGTCTGCGCGCTCCCGCTGTTCACAGCGCTTTCGACATCGCTTCGCTGCTGTC 176  
Db 769 CGTGATCTCTGTGACAGCTCCCGTGTTCACATATACTTCCCTCACTCGGTGATGTCG 828  
QY 177 CGCCACCGGTGTGCGGCGCAGCGTTCGAGTCGCGGAGCGGAGCTTCGCGCGAGCTGAT 236  
Db 829 CCGGCTCGGTGTGCGCGCGCGCATCTGATCATGACAGAGTTCGAGATCGTGCCTGAT 888  
QY 237 GCGGCTGATCGAGGCGGAGAGCTCGACCTTCTTGGCGCGGTTCGCGACCACTTACCACCA 296  
Db 889 GGAGCTCGTGCAGCGGTACCGGTGACGATCTTCCCGCGCATGCTGCCCGCATGCTGCGA 948



QY	297	CCTGGTGGGGCCGCGCGGCGGTTCTCCGCGCGAGCCTCGGATCGGCCTGGC	356
Db	949	GATCGCCAGAGCGCCGAGGTGACCGGTACGACCTGTCTCGATCCGGACCATCATGTC	1008
QY	357	CGGGGGCGGGTCTCTCGGCGCCGGGCT---GCCAAGCGAGTTGCAAGAGACCTTCGGGGT	413
Db	1009	GGGTGCGGCCCGCATGGGGAAGGAGCTCGAGGACACCGTCCGAGCCAAGCTGCCAATGC	1068
QY	414	CCGCTGATCGACGCTACGGCAGCACCGAGACCTCGGGGGCGATCACCATGACCGCC	473
Db	1069	CAAGCTCGGACAGGGGTATCGGATGACGGAGCGGCGCCCGTGTGCAATGTGCCCGGC	1128

Search completed: June 17, 2004, 03:21:44  
Job time : 68.6667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:31:09 ; Search time 293.667 Seconds

(without alignments)  
7787.123 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325

Perfect score: 501

Sequence: 1 agttctacacgtcgggcac.....ccgcgtcgaggctcgatgc 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2998549 seqs, 2282253817 residues

Total number of hits satisfying chosen parameters: 5997098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	285	56.9	692	17	US-10-461-194-11 Sequence 11, Appl
2	285	56.9	85692	17	US-10-461-194-1 Sequence 1, Appl
3	283.4	56.6	86941	17	US-10-461-194-2 Sequence 2, Appl
4	280.2	51.9	644	17	US-10-461-194-23 Sequence 22, Appl
5	268.6	51.6	644	17	US-10-461-194-22 Sequence 23, Appl
6	221.8	44.3	405	17	US-10-461-194-32 Sequence 32, Appl
7	184.2	36.8	77536	10	US-09-940-3168-1 Sequence 1, Appl
8	114	22.8	1497	16	US-10-369-493-32037 Sequence 32037, A
9	109.6	21.9	1500	16	US-10-369-493-43198 Sequence 43198, A
10	107.8	21.5	1566	16	US-10-369-493-42307 Sequence 42307, A
11	101.2	20.2	1485	15	US-10-156-761-373 Sequence 373, Appl
12	101.2	20.2	9025608	15	US-10-156-761-1 Sequence 1, Appl
13	98.8	19.7	2520	15	US-10-149-485-1 Sequence 1, Appl
14	96.2	19.2	1455	15	US-10-156-761-1335 Sequence 1335, Ap

c	15	96.2	19.2	9025608	15	US-10-156-761-1	Sequence 1, Appl
	16	92.8	18.5	1509	15	US-10-156-761-1592	Sequence 1592, Ap
	17	90.8	18.1	1542	16	US-10-369-493-44547	Sequence 44547, A
	18	90.2	18.0	1478	16	US-10-260-238-3996	Sequence 3996, Ap
	19	87.8	17.5	1584	13	US-10-282-122A-11150	Sequence 11150, A
	20	87.8	17.5	18876	15	US-10-329-079-42	Sequence 42, Appl
	21	87.8	17.5	61944	15	US-10-329-079-34	Sequence 34, Appl
	22	86.8	17.3	1572	16	US-10-369-493-28305	Sequence 28305, A
	23	86.8	17.3	1572	16	US-10-369-493-31062	Sequence 31062, A
	24	86.6	17.3	1527	16	US-10-369-493-31819	Sequence 31819, A
	25	85.4	17.0	2006	15	US-10-289-757-143	Sequence 143, App
	26	85.4	17.0	2014	15	US-10-289-757-30	Sequence 30, Appl
	27	84.8	16.9	1467	16	US-10-369-493-28196	Sequence 28196, A
	28	84.8	16.9	1467	16	US-10-369-493-30954	Sequence 30954, A
	29	84.2	16.8	1482	16	US-10-369-493-41560	Sequence 41560, A
	30	82.8	16.5	1473	16	US-10-369-493-31547	Sequence 31547, A
	31	82.4	16.4	1366	13	US-10-425-114-25299	Sequence 25299, A
	32	82.2	16.4	2073	15	US-10-289-757-31	Sequence 31, Appl
	33	81.4	16.2	7788	15	US-10-329-079-8	Sequence 8, Appl
	34	81.4	16.2	37360	15	US-10-329-079-6	Sequence 6, Appl
	35	81.2	16.2	1548	16	US-10-369-493-28210	Sequence 28210, A
	36	81.2	16.2	1548	16	US-10-369-493-30969	Sequence 30969, A
	37	80.6	16.1	2006	15	US-10-182-113A-2	Sequence 2, Appl
	38	80.6	16.1	2075	13	US-10-425-114-889	Sequence 889, App
	39	80.6	16.1	2122	15	US-10-361-460-20	Sequence 20, Appl
	40	79.4	15.8	1527	16	US-10-369-493-35934	Sequence 35934, A
	41	78.6	15.7	1855	15	US-10-289-757-27	Sequence 27, Appl
	42	78.6	15.7	1470	16	US-10-369-493-32475	Sequence 32475, A
	43	77.4	15.4	2039	15	US-10-289-757-142	Sequence 142, App
	44	77.4	15.4	2041	15	US-10-289-757-28	Sequence 28, Appl
	45	77	15.4	1569	16	US-10-369-493-43103	Sequence 43103, A

#### ALIGNMENTS

#### RESULT 1

US-10-461-194-11  
; Sequence 11, Application US/10461194  
; Publication No. US20040077058A1  
; GENERAL INFORMATION:  
; APPLICANT: Hutchinson, Richard C.  
; APPLICANT: Reid, Ralph C.  
; APPLICANT: Hu, Zhihao  
; APPLICANT: Raasher, Andreas  
; APPLICANT: Schirmer, Andreas  
; APPLICANT: McDaniel, Robert  
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
; TITLE OF INVENTION: PRO-GLUTAMINYL PRODUCING POLYPEPTIDE SYNTHASES AND  
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF  
; FILE REFERENCE: 300622009700  
; CURRENT APPLICATION NUMBER: US/10/461,194  
; CURRENT FILING DATE: 2003-06-13  
; PRIOR FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/389,255  
; PRIOR FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/393,929  
; PRIOR FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: US 60/395,275  
; PRIOR FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: US 10/212,962  
; PRIOR FILING DATE: 2002-08-05  
; PRIOR APPLICATION NUMBER: US 60/415,326  
; PRIOR FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: US 60/420,820  
; PRIOR FILING DATE: 2002-10-24  
; PRIOR APPLICATION NUMBER: US 60/433,130  
; PRIOR FILING DATE: 2002-12-13  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus



```

; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/415,326
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/420,820
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/433,130
; PRIOR FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 86941
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-461-194-2

Query Match      56.6%; Score 283.4; DB 17; Length 86941;
Best Local Similarity 72.9%; Pred. No. 3.2e-59;
Matches 365; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 1 ATGTTCTACAGTCGGGACGACGCGGGGCGCCCAAGGGCGTTCGTCACGACGCGCAAC 60
Db |||||
Qy 28184 ATGCTCTACACCTCGGGACACGCGGCGCTGCCGAAAGCGCGTGTCCACCGCGGAAC 28243
Db |||||
Qy 61 TGCCTCTGGTCCGTCGCTTCTCTAGCTGCGCTTCCCGGGTTCGCGACGAGACCGG 120
Db |||||
Qy 28244 TGCTGTGGTCCCTGGCGGCTGTCTAGTACCGGTGACGGGCTGACCGCGAGGACGGT 28303
Db |||||
Qy 121 GTGCTCTGGCGGCTCCCGCTGTTCCACAGCTTTGCGACATCGCTGCTGCTCGCC 180
Db |||||
Qy 28304 GTGCTGTGGCGGCTCGCGCTGTTCCACAGCTTCGACACATCGTGTCTGCTGGCGGCC 28363
Db |||||
Qy 181 ACCGTGTGCGGGCCAGCTCGGATCGCAGCGGAGCTCCGCGAGCGTGTGATGCGG 240
Db |||||
Qy 28364 ACCGCGGTGCGGGCCGCGACCGGATCGTACCGGGGTGTGACCTCCGATGTCTGAC 28423
Db |||||
Qy 241 CTGATCGAGGCGGAGAGCTCGACCTTCTGCGCGGCGTTCGCGACCATCTACCAACCTG 300
Db |||||
Qy 28424 GCGCTGCGCGAGGAGCGTTCGACCTTCATCGCGGAGTGCAGCGCTCTACCAACCTG 28483
Db |||||
Qy 301 GTGCGGCGCGCGCGCGCGGCTTCGCGCGCGAGCTTCGCGGATCGCGCTGCGCGGG 360
Db |||||
Qy 28484 ATCGAGGCTGCGCGCGAGCGAGCTTCGCCACGCGCGAGCTGCGGATCGCGCTGTGGC 28543
Db |||||
Qy 361 GCGCGGCTTCGCGCGCGGCTCGAGCGAGTTCGAGAGACTTCGCGGCTCCCGCTG 420
Db |||||
Qy 28544 GGGCGGTGGCCACGCGCGGACCTGTCTAGTCTGTTGAGGGCCACTTCGAGTGCCTC 28603
Db |||||
Qy 421 ATCGACGCTACGGGACGACCGAGACCTTCGCGGGGCGATCACCATGAACCGCGCGAGCGC 480
Db |||||
Qy 28604 GTCGACGCTACGGATCCACCGAGACCTGTGCGCGGATCGCGGTGAACTGGCCACCGCGC 28663
Db |||||
Qy 481 GCGCGGTGAGGGCTCTGTC 501
Db |||||
Qy 28664 CCACGGGTGAGGGGTCTGTC 28684
Db |||||

RESULT 4
US-10-461-194-23
; Sequence 23, Application US/10461194
; Publication No. US20040077058A1
; GENERAL INFORMATION:
; APPLICANT: Hutchinson, Richard C.
; APPLICANT: Reid, Ralph C.
; APPLICANT: Hu, Zhihao
; APPLICANT: Rascher, Andreas
; APPLICANT: Schirmer, Andreas
; APPLICANT: McDaniel, Robert
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
; TITLE OF INVENTION: PRO-GLUTAMINYLIN PRODUCING POLYKETIDE SYNTHASES AND
; FILE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF
; FILE REFERENCE: 3062200700
; CURRENT APPLICATION NUMBER: US/10/461,194

```

```

; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,255
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/393,929
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/395,275
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/415,326
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/420,820
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/433,130
; PRIOR FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 644
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-461-194-23

Query Match      51.9%; Score 260.2; DB 17; Length 644;
Best Local Similarity 71.8%; Pred. No. 2.5e-53;
Matches 354; Conservative 0; Mismatches 138; Indels 1; Gaps 1;

Qy 9 CAGTCGGGACGACGACGCGGGCGCCCAAGGGCGTTCGTCACGACGCGCAACTGCTCTG 68
Db |||||
Qy 1 CGCTCGGGGACACGCGGCGCTGCCGAAAGGGTGTCTCCACCGCGGAACTGCTCTG 60
Db |||||
Qy 69 GTCCGCTCGCTTCTCTGCTACGTGCGCTTCCCGGGTTCGCGACGAGGACCGGGTCTCTG 128
Db |||||
Qy 61 GTCCCTCGGCGCTCTGCTACGTACCGGTGACGGGCTGACCGCGGAGACCGTGTCTGTG 120
Db |||||
Qy 129 GCGGCTCGGCTGTTCCACAGCTTTCGACATCGCTGCTCTGTCGCGACCGTGTG 188
Db |||||
Qy 121 GCGGCTCGGCTGTTCCCGGAGCTTCGCAATCGTGTCTGTCGCGGCGCCACCGCGGT 180
Db |||||
Qy 189 CGGGGCGGAGGTCGCGGATCGCGGACGCGAGCTCCGCGGACGAGCTGATCGCGCTGATCGA 248
Db |||||
Qy 181 CGGGGCGGCGGACCGGAGTCG--TGACGGGGTGTGACCTCCGATGTCTGAGACGGCTGCG 239
Db |||||
Qy 249 GCGGAGAGCTTCGACCTTCTCTGCGCGGCTTCGCGGACCGACCTACCACTGTTGTCGCGGC 308
Db |||||
Qy 240 CGAGGAGCGGTGCGACCTTCATCGCGGAGTGCAGCGCTTACCACTGTTGTCGAGGC 299
Db |||||
Qy 309 GCGCGGAGCGCGGTTCTCGCGGAGCTTCGCGGATCGGCTTGGCGGGGGCGGT 368
Db |||||
Qy 300 TGCCTGCGAGCGGACTTCGCGCACACCGAGCTCGGATCGCTCTGTCGCGGGCGGT 359
Db |||||
Qy 369 CCTCGGCGCGGGCTCGGAGCGAGTTCGAGAGACCTTCGCGGTTCCGCTGATCGAGCGC 428
Db |||||
Qy 360 GCGCAGCGCGGACCTGTGAGTCTGAGGCGGACCTTCGAGAGTGCCTCTGCTGAGCGC 419
Db |||||
Qy 429 CTACGGGACGACCGGAGCTTCGCGGGGCGATCACCATGAACCGCGGACCGCGCGCGGT 488
Db |||||
Qy 420 CTACGGATCCACCGAGACCTGTGCGCGGATCGCTGCGGCTGCGCGCGCGCGCGGT 479
Db |||||
Qy 489 CGAGGGCTCTGTC 501
Db |||||
Qy 480 CGAGGGGTCTATGC 492
Db |||||

RESULT 5
US-10-461-194-22
; Sequence 22, Application US/10461194
; Publication No. US20040077058A1
; GENERAL INFORMATION:
; APPLICANT: Hutchinson, Richard C.
; APPLICANT: Reid, Ralph C.
; APPLICANT: Hu, Zhihao
; APPLICANT: Rascher, Andreas

```

```

; APPLICANT: Schirmer, Andreas
; APPLICANT: McDaniel, Robert
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
; TITLE OF INVENTION: PRO-GLISDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF
; FILE REFERENCE: 300622009700
; CURRENT APPLICATION NUMBER: US/10/461,194
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,255
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/393,929
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/395,275
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/415,326
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/420,820
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/433,130
; PRIOR FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 644
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-461-194-22

```

Query Match	51.6%	Score	258.6	DB	17	Length	644
Best Local Similarity	71.6%	Pred.	No. 6e-53				
Matches	353	Conservative	0	Mismatches	139	Indels	1
Qy	9	CACGTGGGACACACCGGGCGGCCAAGGGCGTGTCCACGACGACGCAACTGCCCTCTG	68				
Db	1	CACCTGGGACACACCGGGCCTTACCGAAGGGCGCTGTCCACCAGCGGAACTGCTCTGT-59					
Qy	69	GTCCGTGCGCTTCTGTCTACGTGCGCTTCCCGGGGTGTTCGGACACAGGACCGGGTGCCTG	128				
Db	60	GTCCCTGGCGCCTGCTACGTACCGTGCAGGGGACTGACCGCCGAGGACCGTGTGCTGTG	119				
Qy	129	GCGCCTCCCGCTGTTCACAGCCTTTTCGCATCATCGCTGCGTCTCTTCGCGCAACCGTGT	188				
Db	120	GCGCCTCCCGCTGTTCACAGCCTCTCGCATCATGCTGTGTCTGTGGCGGCAACACCGT	179				
Qy	189	CGGGCCAGCGTCCGGATCGCGACGACGAGCTCCGCGACACGTGATGCGGCTGATCGA	248				
Db	180	CGGGCCCGGCACCCGGAATCGTGAACGGGTGTGCAACCCCGATGTCTGAGCGGCTGCG	239				
Qy	249	GGCGGAGAGCTGCACCTTCTTGGCGCGCGTCCGACACCACTTACCACCACTTGGTTCGGGC	308				
Db	240	CGAGGAGGCTCGACCTTATCGCGGAGTSCCGAGGCTTACCACCATCTGATCGAGGC	299				
Qy	309	CGCCCGCAGCGCGGTTCCTCCGCGCGAGCCTTGGGATCGGCTTGGCGCGGGGGCGCGGT	368				
Db	300	TGCCCGGAGGCGGACTTCGCGACGCGCCGAGCTGCGCATCGCTGCTGGGCGGGCGGT	359				
Qy	369	CCTCGGGCGCGGCTGCGAAGCGAGTTCGAGAGACCTTCGGGGTCCGCTGATCGAGCG	428				
Db	360	GGCCACCGGCCGACCTTGTCTGAGGCGCACCTTCGAGTGTCCATCTGTCGACGC	419				
Qy	429	CTAGCGGACGACCGAGACCTTGGGGGCGATCACAATGAACCCGCGGAGCGGGCGCCGCT	488				
Db	420	CTACGGATCCACCGAGACCTGTGGCGCATCGCGTGAATGGCCCAACCGGCCACCGGT	479				
Qy	489	CGAGGGCTCGTGC	501				
Db	480	CGAGGGCTCGTGC	492				

RESULT 6  
US-10-461-194-32/c

```
; Sequence 32, Application US/10461194
; Publication No. U520040077058A1
; GENERAL INFORMATION:
; APPLICANT: Hutchinsonson, Richard C.
; APPLICANT: Reid, Ralph C.
; APPLICANT: Hu, Zhihao
; APPLICANT: Rascher, Andreas
; APPLICANT: Schirmer, Andreas
; APPLICANT: Mcdaniel, Robert
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
; TITLE OF INVENTION: PRO-GLDANAMYCN PRODUCING POLYKETIDE SYNTHASES AND
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF
; FILE REFERENCE: 300622009700
; CURRENT APPLICATION NUMBER: US/10/461,194
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,255
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/393,929
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/395,275
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/415,326
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/420,820
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/433,130
; PRIOR FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-461-194-32

Query Match      44.3%; Score 221.8; DB 17; Length 405;
Best Local Similarity 72.1%; Pred.No. 4.1e-44;
Matches 289; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy   34  AAGGGCGTGTGTCACGCAGCGCAACTGGCTCTGTGGTCGGTGCGTCTCTGTCTGTCTGTCCG 93
Db   405  AAAGCGTGCTGTTCACCAGACGGAACTGCTGTGGTCCCCTGGCGCCTGCTACGTACCG 346

Qy   94  TTCGCCGGTTGTGGACACAG3ACCGGGTGTCTTGGCCGTCTCCCGCTGTTCACACAGCTT 153
Db   345  GTACGGGGCTGACCGCGAG3ACCGTGTCTGTGGCCGTCTGCGCTGTTCACAGCTC 286

Qy   154  TCGCACATCGCTTGGTCCTGGCCACCGTGTGTGGGGCCACGCTCCGGATCGCCGAC 213
Db   285  TCGCACATCTGTGTCTGTCTGGCGCCACCGCTGTGGGGCCGACCCGGATCGTGAC 226

Qy   214  GCGAGCTCCGCCGACGACGTGATCGCGCTGATCGAGCGGAGAGCTCGACTTCTCTGCC 273
Db   225  GGGGTGTGCGACCTCCGATGTGTGACGCGCTGTGGCGAGGAGCGTTCGACTTCATCGCC 166

Qy   274  GGCGTCGGACCATTCACCACTGGTGGGGCCCGCCGCGAGCGGGTTCCTCCGCG 333
Db   165  GGAATGCGGACGCTCTACCACTCTGATCGAGGTCTGCGCGAGCGGCACTTCGCCACG 106

Qy   334  CCAGAGCTCGGGATCGCGCTTGGCGGGGGCGCGTCTCTCGCGCGCGGCTCGGAAGCGAG 393
Db   105  CCCGAGCTCGGATCGCGCTCGTGTGGCGGGCGGTGGCCAAGCGGCACTGTCTAGGTCTG 46

Qy   394  TTGGAAGAGACTTCGGGGTCCCCCTGATCGAACGCTTACGG 434
Db   45    TTCAGAGCGCACTTCGGAGTGTCCACTCGTTCGACGCTTACGG 5
```

RESULT 7  
US-09-940-316B-1/c  
; Sequence 1, Application US/09940316B





NAME/KEY: gene  
LOCATION: (124)..(987)  
OTHER INFORMATION: ech  
FEATURE:  
NAME/KEY: RBS  
LOCATION: (978)..(981)  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (987)..(2459)  
OTHER INFORMATION: fcs, gene of feruloyl-CoA synthetase  
FEATURE:  
NAME/KEY: gene  
LOCATION: (987)..(2462)  
OTHER INFORMATION: fcs  
US-10-149-485-1

Query Match 19.7%; Score 98.8; DB 15; Length 2520;  
Best Local Similarity 54.5%; Pred. No. 8.7e-15;  
Matches 243; Conservative 0; Mismatches 197; Indels 6; Gaps 2;

QY 2 TGTCTTACACGTCGGGACGACCGGGCGGCCCAAGGGCGTGTGTCCAGCGACGCGCAACT 61  
Db 1438 TGTGTATACACTCGGCGAGCACCGGACGCGCCAGGGCGGATGCTCACCCAGGCAACC 1497

QY 62 GCTCTGGTCCGTCGCTTCTCTGTCTACGTGCGGTTCCCGGGTGTTCGACCGAGACCGGG 121  
Db 1498 TCACCTGGAACTCGCTCAACGTCCTCTGTGGAGA---CCGACCTGGCGAGCGACGAGCGGG 1554

QY 122 TGTCTGGCGCTCCCGCTGTTCACAGCTTTTCGACATCGCTGCTCTCTCCGCCA 181  
Db 1555 CACTGTGTCGGCGCGCTGTTCACGCGCCGCTGTTCGACATCGCTGCTCTCTCCGCCA 1614

QY 182 CGTGTGTCGGGCGACGCTCGGATCGCGACCGGCGAGCTTCGCGACGAGCTGATCGGC 241  
Db 1615 TGCTCAAGGGCGGACGCTGATCTTGACCTCCGCTTCGACCCCGCGCTGCTGTCCG 1674

QY 242 TGATCGAGCGGAGAGCTGACCTTCTGCGCGGCTGCGGACCTTACACCACTGCG 301  
Db 1675 CGGTGGAAACGAGAGCGGTCACGCTGCTGTTCGCGCTGCGCCACGATGTACCGCGATCG 1734

QY 302 TGGCGCGCGCGGCGAGCGGTTCTCCGCGCGAGCTTCGCGAGCTTCGCGAGCTTCGCGGG 361  
Db 1735 CGCGCACCCCGCGGTGCGCGAGCGGACCTGTCAGCTTCGCGAGCTTCGCGGG 1794

QY 362 GCGCGGTCTTCGCGCGCGGCTGCGAGCGAGTTTCAAGAGACCTTCGCGGTTCCCGCTGA 421  
Db 1795 GCGCGCGGTTCGCGCGGACCTTCGCGAGCGCTACCTCGA---CCGCGGCTCGCGTTG 1851

QY 422 TCAGCGCTACGGGAGCGACCGAGCC 447  
Db 1852 TGCAGGCGCTACGGCATGACCGAGGCC 1877

## RESULT 14

US-10-156-761-1335  
Sequence 1335, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10156.761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-05-30

PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 1335  
LENGTH: 1455  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1455)  
US-10-156-761-1335

Query Match 19.2%; Score 96.2; DB 15; Length 1455;  
Best Local Similarity 51.3%; Pred. No. 4e-14;  
Matches 249; Conservative 0; Mismatches 233; Indels 3; Gaps 1;

QY 2 TGTCTTACACGTCGGGCGGACCGGGCGGCCCAAGGGCGTGTGTCCAGCGACGCGCAACT 61  
Db 416 TGTATCTACACGAGCGGACCGGGCGGCCCAAGGGTGTGTCTCGACCAACGCCA-- 473

QY 62 GCTCTGGTCCGTCGCTTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 121  
Db 474 -CTCGCGCCCATGGCGGAGATGATACCGGCAACCGCCCGGCTGACCGAGACCGACCA 532

QY 122 TGTCTGGCGCTCCCGCTGTTCACAGCTTTTCGACATCGCTGCTGCTGCTGCTGCTGCTGCT 181  
Db 533 GCTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 592

QY 182 CGTGTGTCGGGCGACGCTTCGGATCGCGAGCGGAGCTTCGCGGACGAGCTGATGCGGC 241  
Db 593 TGTGTGCGCGCGCGCGCTTCAGGTTGGCGGGCGGTTCCGGGCGGAGACCTTCTTCGACC 652

QY 242 TGATCGAGCGGAGAGCTGACCTTCTGCGCGGCTTCGCGAGCTTCGCGAGCTTCGCGAGCTGG 301  
Db 653 TGTGTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 712

QY 302 TGGCGCGCGCGGCGAGCGGTTTCTCGCGCGCGGAGCTTCGCGAGCTTCGCGAGCTTCGCGGG 361  
Db 713 CGGAATGCGCGGACGAGCTACGCGCGGACGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 772

QY 362 GCGCGGTCTTCGCGCGCGGCTGCGAGGAGCTTCGAGAGAGCTTCGCGGGTTCGCGGTGA 421  
Db 773 GCGCGCGGATGCGCGCGGCTCATCGAGCGGTTTCGAGCGCGCTACGACATCCCGTCC 832

QY 422 TCAGCGCTACGCGACGACCGGACCTTCGCGGCGGATGACCATGACCGCGGAGCGG 481  
Db 833 TGAAGGCGTACGCGCTTCTCGAGGGGACCTTCGCGGCTTCACACCAACCGCTGTACGGCC 892

QY 482 CCGCG 486  
Db 893 GCGCG 897

## RESULT 15

US-10-156-761-1/c  
Sequence 1, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10156.761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109

```
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      19.2%; Score 96.2; DB 15; Length 9025608;
Best Local Similarity 51.3%; Pred. No. 1e-14;
Matches 249; Conservative 0; Mismatches 233; Indels 3; Gaps 1;

Qy 2 TGTTCACGTCGGGACGACCGGGGCGCCCAAGGGGCTGTGTCTCACGCGAGCGCAACT 61
Db |||||
Db 1660786 TGATCTACACAGCGGAGCAGCACGGGGCGCCCAAGGGTGTGTCTCGACACGCCAA-- 1660729
Qy 62 GCCTCTGGTCCGTCGCTTCTCTGCTAGCTGCGGTTCCCGGGTGTGCGACCCAGGACCGGG 121
Db |||||
Db 1660728 -CCTCGCGCCATGGCGAGATGATGACCGGACCGCCCGGCTACCGAGACCGACCA 1660670
Qy 122 TGCTCTGCGCCGCTCCCGCTGTTCACAGCCCTTTCGCAATCGCCTGGCTCCTGTCCGCCA 181
Db |||||
Db 1660669 GCCTGCTGATCTCGCGCTGTTCACGCTCAACGGGATCGTGTGCGGGTGTGTCCCGC 1660610
Qy 182 CCGTGGTCGGGGCCAGCGTCCGGATCGCGACGGCAGCTCCGCCGACGACGATGCGGC 241
Db |||||
Db 1660609 TGCTGGCGGGCGCGCGGTCAACGGTGGCGGGCGGTTCCGGGCGGAGACCTTCTTCGACC 1660550
Qy 242 TGATCGAGGGCGGAGAGCTCGACCTTCTTGGCGCGGTGCGGACCACTTACCACACCTGG 301
Db |||||
Db 1660549 TGGTGGCACCGGTCCGCCCGACCTGCTTCTCCGGGTCCCGGCGATCTACTCGATGCTCG 1660490
Qy 302 TGGCGGGCGCGCGGAGCGCGGTTTCTCCGGCGCGAGCTTCGGATCGGCTGGCCGGGG 361
Db |||||
Db 1660489 CCGAATCGCCGACCAACGTCAGCGCCGACACTCTGTCGTTACGGTTCCGCGCTGCGGG 1660430
Qy 362 GCGCGGTCTCGCGCGCGCGGTGCGAGCGAGTTCGAAGAGACCTTCGGGGTCCCGCTGA 421
Db |||||
Db 1660429 CGGCGCCGATGCCGCGCGCGCTCATCGAGCGGTTTCGAGCGCGCTTACGACATCCCGGTCC 1660370
Qy 422 TCGAGCGCTACGCGAGCACCGAGACCTCGGGGGGATCACCATGAACCCGCGGACGCGG 481
Db |||||
Db 1660369 TGGAGGGCTACGSCCTCTCCGAGGGGACCTCGCGCTCCACCAACCCCGCTGTACGGCC 1660310
Qy 482 CCCGC 486
Db |||||
Db 1660309 GGCGC 1660305
```

Search completed: June 17, 2004, 03:36:44  
Job time : 318.667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:18:23 ; Search time 1960 Seconds  
(without alignments)  
7633.139 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325

Perfect score: 501

Sequence: 1 atgttacaactcgagcac.....ccgcgtcgaggctcgtgc 501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_esti:\*

9: gb\_esti:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

.Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	18.0	684	13	CA012845
2	84.8	16.9	564	14	CA012845 HT06J23r
3	83.6	16.7	666	12	CA012845 HT06J23r
4	80.8	16.1	588	13	CA012845 HT06J23r

5	80.6	16.1	2090	11	AY105108
6	80.4	16.0	543	9	AI665488
7	75.4	15.0	536	14	CD056352
8	74.4	14.9	661	13	BQ743361
9	74	14.8	569	13	BQ789501
10	72.2	14.4	830	29	CG34022
11	72.2	14.4	876	29	CG310929
12	71.2	14.2	706	13	CA143501
13	71	14.2	598	10	BE426294
14	71	14.2	982	13	CA143501
15	70.4	14.1	668	12	BU481484
16	70.2	14.0	563	29	CG679053
17	70.2	14.0	564	29	CG679056
18	69.8	13.9	744	14	CA270511
19	69.8	13.9	925	29	CNS0091P
20	69.6	13.9	616	13	BU100715
21	69.4	13.9	570	14	CA549499
22	69.2	13.8	1139	28	BZ552857
23	68.8	13.7	875	12	BG344702
24	68.8	13.6	935	29	CNS0066X
25	68.2	13.6	629	12	BI959998
26	67.2	13.4	476	10	BE215262
27	67.2	13.4	709	14	CD433365
28	67	13.4	935	29	CNS0066X
29	66.8	13.3	925	29	CNS0091P
30	66.8	13.3	982	13	CA143501
31	66.2	13.2	1315	11	AY106966
32	65	13.0	438	10	BE597495
33	63.8	12.7	602	13	BU998998
34	63.8	12.7	742	29	CG297240
35	63.8	12.7	1070	28	BZ557810
36	63.6	12.7	2079	11	AY108238
37	63.2	12.6	600	28	BZ895297
38	63.2	12.6	703	14	CA196875
39	63	12.6	642	13	CA025271
40	62.8	12.5	1101	29	CNS01751
41	62.8	12.5	726	10	BE519742
42	62.6	12.5	703	14	CA162305
43	62	12.4	643	13	CA162305
44	62	12.4	646	13	CA162305
45	62	12.4	932	29	CNS0072Q

## ALIGNMENTS

RESULT 1  
CA012845  
LOCUS  
DEFINITION HT06J23r HT Hordeum vulgare subsp. vulgare cDNA clone HT06J23  
684 bp mRNA linear EST 23-OCT-2002  
5-PRIME, mRNA sequence.  
ACCESSION CA012845  
VERSION CA012845.1 GI:24290189  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooideae; Triticeae; Hordeum.  
1 (bases 1 to 684)  
Zhang, H., Potokina, E., Michalek, W., Weschke, W., Stein, N. and Graner, A.  
Barley ESTs from germinating seeds  
Unpublished (2002)  
Contact: Stein Nils  
Molecular Markers Group, Department Genbank  
Institute of Plant Genetics and Crop Plant Research (IPK)  
Corrensstr. 3, 06466, Gatersleben, Germany  
Tel: 039482-5522  
Fax: 039482-5595  
Email: stein@ipk-gatersleben.de  
Insert Length: 684 Std Error: 0.00  
Plate: 6 row: J column: 23

```

FEATURES
  source
    Seq primer: M13rev.
    Location/Qualifiers
      1..684
        /organism="Hordeum vulgare subsp. vulgare"
        /mol_type="mRNA"
        /cultivar="barke"
        /sub_species="vulgare"
        /db_xref="GABI:261368"
        /db_xref="taxon:112509"
        /clone="HT06J23"
        /tissue_type="endosperm early"
        /dev_stage="0-16 hours after imbibition"
        /lab_host="XL10-Gold"
        /clone_lib="HT"
    /note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

ORIGIN
  Query Match      18.0%; Score 90; DB 13; Length 684;
  Best Local Similarity 52.3%; Pred. No. 0.0018;
  Matches 253; Conservative 0; Mismatches 220; Indels 11; Gaps 2;

QY 6 CTACAGCTGGGACACACCGGGCGGCCCAAGGCGGTGTCCTCCACGACGCGAATGCTT 65
DB 201 CTACTCGTGGGACACACCGGGCTCCCAAGGCGGTGTCCTCCACGACGCGAATGCTT 260
QY 66 CTGGTCCGTGCG-----TTCTGTAGTGGTCCGTTCCCGGGTTGTCGACACGAC 117
DB 261 CACACGCTGCGCCACGACGCTGACGCGGGAGAACCCGAACTGTACTTCGGGGAGGAGAC 320
QY 118 CGGGTCTCTGGCGCGCTCCCGCTGTTCCACAGCTTTCCGACATCGCTCGCTGCTGTC 177
DB 321 GTGGTGTCTGGCTGCTGCGTGTTCACATCTACTCTGCTCACTCGCTGCTGCTGGCG 380
QY 178 GCCACGCTGGTGGGCGCCAGCGTCCGATGCGCGACGCGAGCTCCCGCGACGACGTGATG 237
DB 381 GGGCTCGGGCGCGGTGCGGATCGTGTATCATGCGCAAGTTTCACACGCGGTGCTGGTG 440
QY 238 CGGCTGATCAGCGCAGAGCTGACCTTCTGCGCGGGTGGCGGACCACTACACAC 297
DB 441 CGCCTGTGGCGCGCACGCGGGTCACTGCGGCGCTTCTGTCGCGCCCTCGTGTGCGAG 500
QY 298 CTGGTGGCGCGCGCGCGCGCGGTTCCTCGCGCGCGAGCTTGGGATCGGCTCGGCG 357
DB 501 ATGCCAAGACGACCGGGTACGCGCGCGCGACCTCGCCCTCATCGCATGCTCATGTCC 560
QY 358 GGGGGCGCGGTCTCTGGCGCGCGGGTGGCGAAGCGAGTTC---GAAGAGACCTTGGGGTTC 414
DB 561 GCGCGCGCGCGCATGCGCAAGGACCTCCAGAGCGCTTCATGCGCAAGATCCCGCAAGCC 620
QY 415 CCCTGATCAGCGCTACGCGACGACCGACGACCTGCGGGCGGATCACCATGACCGCGCG 474
DB 621 GTCTCGGACAGGGGTACCGGATGACGAGGCGCGCGCGCGTGTGTCTCATGTGCTGGCG 680
QY 475 GACG 478
DB 681 TTCG 684

RESULT 2
CD865561
LOCUS
DEFINITION
  AZO2:101D12F001107 AZO2 Triticum aestivum cDNA clone AZO2101D12, mRNA sequence.
ACCESSION
  CD865561
VERSION
  CD865561.1 GI:32549377
KEYWORDS
  EST.

SOURCE
  Triticum aestivum (bread wheat)
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
  REFERENCE
    1 (bases 1 to 564)
  AUTHORS
    Genoplante.
  TITLE
    Genoplante, a major partnership french program in plant genomics
  JOURNAL
    Unpublished (2003)
  COMMENT
    Contact: Genoplante
    Genoplante
    93, rue Henri Rochefort 91025 EVRY CEDEX France
    Tel: 33 1 69 47 54 00
    Fax: 33 1 69 47 54 10
    This sequence has been generated in the framework of the french plant genomics programme 'Genoplante' (http://www.genoplante.com and http://genoplante-info.infobiogen.fr).
  FEATURES
    source
      1..564
        /organism="Triticum aestivum"
        /mol_type="mRNA"
        /cultivar="recital"
        /db_xref="taxon:4565"
        /clone="AZO2101D12"
        /tissue_type="root"
        /clone_lib="AZO2"
  ORIGIN
    Query Match      16.9%; Score 84.8; DB 14; Length 564;
    Best Local Similarity 53.4%; Pred. No. 0.0094;
    Matches 251; Conservative 0; Mismatches 207; Indels 12; Gaps 3;

QY 8 ACACGTCGGGACGACGCGGGCGGCCCAAGGCGGTGTCCTCCACGACGCGAATGCTTCT 67
DB 98 ACACCTCGCGACGACGACGCGGCCCAAGGCGGTGTCCTCCACGACGCGAATGCTTCT 157
QY 68 GGTCTGCTGCTTCTCTGCTACGTGCGCTTCCCGGGTGTTCGACACGACGCGGTGCTCT 127
DB 158 CCTCGTCCAGAACATCCGCGCGGTGTACG---GGTTTCGCGAGACGACGCGCGTGG 214
QY 128 GCGCGTCCCGCTGTTTCCACAGCTTTTCACATCGCTTCCGCTGCTGCTGCGCACCGTGG 187
DB 215 TGACACTGCGCTGTTTCCACGCTGACGCGCTCATGTGCGCGTGTCTCTCTGCGCT 274
QY 188 TCGGGGCCAGCTTCGCGATCGCGACGCGCAGC-----TCGCGCAGACGTGATGCGGC 241
DB 275 CCGCGCGTGGTGGCGCTCCCGCGCGCGCGCTTCTCGGGTCCAGCTTCTCGGGTCCAG 334
QY 242 ---TGATCGAGGCGGAGAGCTGACCTTCTGCGCGCGGTGCGACCACTACACCAACC 298
DB 335 ACATGCTCGCGCGCGCGCGCTGCTGTACACGCGGTGCGCCACGATCCACAGATCATCC 394
QY 299 TGTGCGGGCGCGCGCGCGCGGTTCCTCGCGCGCGAGCTTCCGCGCGCGCTGCGGATCGGCG 358
DB 395 TGGACCGGACGCGCTGCTGCGCGCGCGGTACCCGCGCGCTGCGGTTCTGTCGCGAGCT 454
QY 359 GGGCGCGCTTCTCGCGCGCGGCTCGGAAGCGAGTTTCAAGAGACCTTCGCGGGTCCCGC 418
DB 455 GCACCGCGTCTGCGCGCGCGGTGATCTGAGCGCGCTGAGGCGGGTTCGGGGCGCGG 514
QY 419 TGATCGACGCTACGCGCAGCACCGAGACCTTGGCGGCGGATCACCATGAAC 468
DB 515 TGCTGGAGGCGGTACCCCATGACGAGAGCGCTCCACCATGATGACCTCCAAC 564

RESULT 3
BI958356
LOCUS
DEFINITION
  BI958356 666 bp mRNA linear EST 22-OCT-2001
  HVSMEN0014124f Hordeum vulgare rachis EST library HVCDNA0015
  (normal) Hordeum vulgare subsp. vulgare cDNA clone HVSMEN0014124f, mRNA sequence.
ACCESSION
  BI958356
VERSION
  BI958356.1 GI:16309611

```









RESULT 9	BQ789501	569 bp	mRNA	linear	EST 26-JUL-2002
LOCUS	WHE4162_B02	D04ZS	Wheat CS whole plant cDNA library	Triticum	
DEFINITION	aeestivum cDNA clone WHE4162_B02_D04,		mrna sequence.		
ACCESSION	BQ789501				
VERSION	BQ789501.1	GI:21997973			
KEYWORDS	EST.				

**FEATURES**  
**SOURCE**

```

/db xref="taxon:4565"
/clone="WHE4162_B02_D04"
/tissue_type="Roots, leaves, crown, stem and sheath"
/dev_stage="Adult"
/lab host="E. coli SOLR"
/clone lib="Wheat CS whole plant cDNA library"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid
pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Plant
tissues from wheat cv. CS grown to full tillering stage in
greenhouse were collected at University of California,
Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total
RNA was prepared from leaves (young leaf and third leaf),
whole roots, crown, stem and sheath tissues, and then
equal quantities of RNA were pooled from these
samples. PolyA was purified from the pooled RNA, a cDNA
library was made, and the cDNA clones were in vivo excised
to give plusescript SK(-) phagemids in J. Dvorak's lab (E.
Akhunov, J. Dvorak) at the University of California,
Davis. Colony plating, plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

```

ORIGIN

Query Match 14.8%; Score 74; DB 13; Length 569;  
 Best Local Similarity 51.6%; Pred. No. 0.32;  
 Matches 224; Conservative 0; Mismatches 200; Indels 10; Gaps 2;

QY 2 TGTCTTACAGCTCGGGCACCACCGGGCGGCCAAGGCGCTGTGTCCACGACGCAACT 61  
 DB 128 TGCCTTACTCTCTCGGCACACACCGGGCTGCCAAGGCGTATGCTCACGACGCGACGC 187  
 QY 62 GCCTCTGGTTCGGTCTTC-----CTGCTACGTGCGCTTCCCGGGTTGTGCGACC 112  
 DB 188 TCATCACAGCGTCCGCCAGCAGGTGACCGGGGAGAACCCGAACCTGTACTTCAGCAAG 247  
 QY 113 AGGACCGGGTGTCTGCGCGCTCCCGCTGTTCACAGCCTTTTCACATCGCCTGCGTCC 172  
 DB 248 AGGACGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 307  
 QY 173 TGTCCGCCACCGTGTGCGGGGCCAGCGTCCGGATCGCGCGAGCTTCGCGCCGACGACG 232  
 DB 308 TGGGGGCGCTGCGCGCGCGGCTGCGGATGCTGATCATGCGCAAGTTGACATCGCGCGCG 367  
 QY 233 TGATCGCGCTGATCGAGCGGAGAGCTCGACTTCTGCGCGCGGCTGCGGACGACCTTACC 292  
 DB 368 TGGTGGAGCTGCTGCGCGCGCACCGGCTACGATCGCGCGCTTCTGCTG-CGCGCATCGTG 426  
 QY 293 ACCACTGTGTCGGCGCGCGCGCGCGCGCGCGCTTCTCGCGCGCGAGCTTCGAGAGACCTTCGGGG 352  
 DB 427 TGGAGATCGCAAGAGCGCGCGCGCGCGCGCGCGCTTCTATGCGCGCGAGTTCAGTCCCA 486  
 QY 353 TGGCGCGCGCGCGCGCTTCTCGCGCGCGCGCGCTTCGAGAGACCTTCGAGAGACCTTCGGGG 412  
 DB 487 TGTCCG 546  
 QY 413 TCCGCGTATCGAC 426  
 DB 547 ACCGCGTGTCTGCG 560

RESULT 10  
 CC634022  
 LOCUS OGULM87TV ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBma0469P05, linear GSS 19-JUN-2003  
 DEFINITION genomic survey sequence.  
 ACCESSION CC634022  
 VERSION CC634022.1 GI:32010601  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 830)  
 AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 TITLE Consortium for Maize Genomics  
 JOURNAL Unpublished (2002)  
 COMMENT Other GSSs: OGULM87TH  
 Contact: Cathy Whitelaw  
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TF  
 Class: sheared ends.  
 Location/Qualifiers  
 1..830  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"

/clone="ZMMBma0469P05"

/clone lib="ZM\_0.7\_1.5\_KB"  
 /note="Vector: pBCSK-; Site:1: HindII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN

Query Match 14.4%; Score 72.2; DB 29; Length 830;  
 Best Local Similarity 51.1%; Pred. No. 0.59;  
 Matches 201; Conservative 0; Mismatches 183; Indels 9; Gaps 1;

QY 2 TGTCTTACAGCTCGGGCACCACCGGGCGGCCAAGGCGCTGTGTCCACGACGCG----- 56  
 DB 324 TGCCTTACTCTCTCGGCACACCGGGCTGCCAAGGCGGTGATGCTCTCGCACCGTGGGC 383  
 QY 57 ----CAACTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 112  
 DB 384 TGGTACACAGCGTGGCGCAGCTGCTCGACGCGGACACCGGAACTCTACTTCGCGGAGG 443  
 QY 113 AGGACCGGGTGTCTGCGCGCTCCCGCTGTTCACAGCCTTTCCACATCGCCTTGGCTGCC 172  
 DB 444 ACAGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 503  
 QY 173 TGTCCGCCACCGTGTGCGGGGCCAGCGTCCGGATCGCGCGAGCTTCGCGCCGACGACG 232  
 DB 504 TGTCCGGGATGCGCGCGCGCGCGCGCGCGCTTCTGATCATGACGCGCTTCGACACTCTCGCA 563  
 QY 233 TGATCGCGCTGATCGAGCGGAGAGCTCGACTTCTTCTGCGCGCGGCTGCGGACGACCTTACC 292  
 DB 564 TGTTCGAGCTGCTGAGAGAGGCA CGGCATACGATCGTGTGCTGCTGCTGCTGCTGCTGCTG 623  
 QY 293 ACCACTGTGTCGGCGCGCGCGCGCGCGCGCTTCTCGCGCGCGAGCTTCGCGGATCGGCG 352  
 DB 624 TGGAGATGCTCAAGAGCGACCGCATCGACCGCGCGCGCGCGCGCTTCTGCTGCTGCTGCTG 683  
 QY 353 TGGCGCGCGCGCGCGCTTCTCGCGCGCGCGCGCTTCGCGCGCGAGCTTCGCGGATCGGCG 385  
 DB 684 TCTCGGGGCG 716

RESULT 11  
 CC310929/c  
 LOCUS OG3DO25TH ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBma0783E01, linear GSS 26-AUG-2003  
 DEFINITION genomic survey sequence.  
 ACCESSION CC310929  
 VERSION CC310929.1 GI:34228089  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 876)  
 AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 TITLE Consortium for Maize Genomics  
 JOURNAL Unpublished (2002)  
 COMMENT Other GSSs: OG3DO25TV  
 Contact: Cathy Whitelaw  
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TR  
 Class: sheared ends.  
 Location/Qualifiers  
 1..876  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"

FEATURES  
 source









score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 22:09:53 ; Search time 2276.33 Seconds  
(without alignments)  
9539.403 Million cell updates/sec

Title: US-10-042-665A-3 COPY 8434 8934

```

Perfect score: 501
Sequence: 1 tcgctttccgctggtcgagct.....cgcctggtcggtcggtg 501

```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 . Gapext 1.0

Searched: 3470272 BEQS, 21671516995 residues

Total number of hits satisfying chosen parameters:	6940544
--	---------

Minimum DB seq length: 0

```
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

GENERAL:

1. *Staphylococcus aureus* (S. aureus) is a Gram-positive, spherical bacterium that is commonly found on the skin and in the nose of humans and animals. It is a facultative anaerobe, meaning it can grow with or without oxygen. S. aureus is known for its ability to cause a wide range of infections, including skin abscesses, boils, and cellulitis. It is also a common cause of food poisoning and can be found in various environments, including hospitals and homes.

z: gb\_lug: \*

**gr - III:**

gr\_011: \* \* \*

5: gm\_ov: +

```
gb_pat:*
```

1: \*pāṇ:\*

\*:gb\_p1:\*

\*9: gb\_pr:\*

10: gb\_ro:\*

```
11: gb_stg:*
```

112: gb\_sy:\*

113: gb\_un:\*

14: gb-vi:\*

15: em\_ba:\*

```
16: em_fun:*
```

```
17: em_hum:*
```

```
18: em_in:*
```

19: em\_mu:\*

20: em om: \*

21: em\_or:★

em\_ov: \*

23. em\_nat.\*

23: em pa: \*

pred. No. is the number of results predicted by chance to have a

RESULT 1	ACCESSION	REFERENCE
AMM223012	VERSION	AUTHORS
LOCUS	KEYWORDS	TITLE
DEFINITION	SOURCE	
	ORGANISM	

AMM223012 53784 bp DNA linear BCT 09-FEB-1998  
Amycolatopsis mediterranei genes encoding rifamycin polyketide  
synthases. ORFs 1 to 5.

## ALIGNMENTS

Result No.	Score	Query		DB	ID	Description
		Match	Length			
C	1	501	100.0	53784	1	AMW223012
	2	501	100.0	53789	6	AG9720
	3	501	100.0	90445	1	AF040570
	4	435.4	86.9	39000	1	AF262754
	5	334.6	66.8	5676	6	AG6718
	6	332	66.3	123580	1	AF263912
	7	332	66.3	125401	6	AX211739
C	8	328.2	65.5	20394	1	SNAX132222
	9	328.2	65.5	20394	6	AXNA132222
	10	328.2	65.5	84985	1	SNA278573
	11	328.2	65.5	104326	1	AB070940
	12	328.2	65.5	320150	1	AP005033
	13	325	64.9	82746	1	AP453501
	14	320.8	64.0	113193	1	AF357202
C	15	320.8	64.0	113193	6	AX211705
	16	317.6	63.4	65140	6	AX211705
	17	317	63.3	78210	1	AB070949
	18	317	63.3	308425	1	AP005022
	19	316	63.1	138203	1	AX130323
	20	314.4	62.8	77534	1	AF235504
	21	314.4	62.8	77536	6	BD235937
C	22	314.4	62.8	77536	6	AX271638
	23	312.2	62.3	107379	1	SHGCP1R
	24	311.8	62.2	290850	1	SC0939127
	25	311.2	62.1	39514	1	SGR300302
	26	311.2	62.1	132544	1	AF251085
	27	309.6	61.8	28732	1	AF082100
	28	309	61.7	107379	1	SHGCP1R
C	29	309	61.7	281450	1	AP005032
	30	307.4	61.4	15872	6	AK362090
	31	307.4	61.4	15872	6	AK362091
	32	307.4	61.4	15872	6	BD217171
	33	307.4	61.4	30000	6	AX250261
	34	307.4	61.4	43280	1	SFU789289
	35	307.4	61.4	103450	1	AF440781
C	36	305.8	61.0	38506	6	BD232534
	37	305.8	61.0	38506	6	AR271680
	38	305.8	61.0	38506	6	AR277664
	39	304.2	60.7	4674	6	BD235956
	40	304.2	60.7	4674	6	AR271657
	41	304.2	60.7	4725	6	BD235955
	42	304.2	60.7	4725	6	BD271656
C	43	304.2	60.7	4737	6	BD235958
	44	304.2	60.7	4737	6	AR271659
	45	304.2	60.7	4767	6	BD235957
	46	304.2	60.7	4767	6	BD235957
	47	304.2	60.7	4767	6	BD235957
	48	304.2	60.7	4767	6	BD235957
	49	304.2	60.7	4767	6	BD235957
C	50	304.2	60.7	4767	6	BD235957
	51	304.2	60.7	4767	6	BD235957
	52	304.2	60.7	4767	6	BD235957
	53	304.2	60.7	4767	6	BD235957
	54	304.2	60.7	4767	6	BD235957
	55	304.2	60.7	4767	6	BD235957
	56	304.2	60.7	4767	6	BD235957
C	57	304.2	60.7	4767	6	BD235957
	58	304.2	60.7	4767	6	BD235957
	59	304.2	60.7	4767	6	BD235957
	60	304.2	60.7	4767	6	BD235957
	61	304.2	60.7	4767	6	BD235957
	62	304.2	60.7	4767	6	BD235957
	63	304.2	60.7	4767	6	BD235957
C	64	304.2	60.7	4767	6	BD235957
	65	304.2	60.7	4767	6	BD235957
	66	304.2	60.7	4767	6	BD235957
	67	304.2	60.7	4767	6	BD235957
	68	304.2	60.7	4767	6	BD235957
	69	304.2	60.7	476		

RESULT 1	AMM223012	53784 bp	DNA	linear	BCT 09-FEB-1998
LOCUS	AMM223012				
DEFINITION	Amycolatopsis mediterranei genes encoding rifamycin polyketide synthases, ORFs 1 to 5.				
ACCESSION	AJ223012				
VERSION	AJ223012.1	GI:2764760			
KEYWORDS	ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.				
SOURCE	Amycolatopsis mediterranei				
ORGANISM	Amycolatopsis mediterranei				
	Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;				
	Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.				
REFERENCE	1				
AUTHORS	Shannon, T., Toupet, C., Engel, N. and Goff, S.				
TITLE	Cloning and sequence analysis of the putative rifamycin polyketide				





RESULT 3  
AF040570  
LOCUS AF040570 90445 bp DNA linear BCT 16-APR-2001  
DEFINITION Amycolatopsis mediterranei rifamycin biosynthetic gene cluster.  
ACCESSION AF040570 AF040571  
VERSION AF040570.2 GI:123331604  
KEYWORDS  
SOURCE Amycolatopsis mediterranei  
ORGANISM Amycolatopsis mediterranei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.  
REFERENCE 1 (bases 1 to 90445)  
AUTHORS August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R., Yu, T.W.,  
Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X., Hutchinson, C.R. and  
Floss, H.G.  
TITLE Biosynthesis of the ansamycin antibiotic rifamycin: deductions from  
the molecular analysis of the rif biosynthetic gene cluster of  
Amycolatopsis mediterranei S699  
JOURNAL Chem. Biol. 5 (2), 69-79 (1998)  
MEDLINE 98174059  
PUBMED 9512878  
REFERENCE 2 (bases 1 to 90445)  
AUTHORS Kim, C.G., Yu, T.W., Fryhle, C.B., Handa, S. and Floss, H.G.  
TITLE 3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the  
formation of the precursor of mC7N units in rifamycin and related  
antibiotics  
JOURNAL J. Biol. Chem. 273 (11), 6030-6040 (1998)  
MEDLINE 98165773  
PUBMED 9497318  
REFERENCE 3 (bases 1 to 90445)  
AUTHORS Yu, T.W., Muller, R., Muller, M., Zhang, X., Draeger, G., Kim, C.G.,  
Leistner, E. and Floss, H.G.  
TITLE Mutational analysis and reconstituted expression of the  
biosynthetic genes involved in the formation of  
3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin  
biosynthesis in Amycolatopsis mediterranei S699  
JOURNAL J. Biol. Chem. 276 (16), 12546-12555 (2001)  
MEDLINE 21201076  
PUBMED 11278540  
REFERENCE 4 (bases 1 to 90445)  
AUTHORS August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R.,  
Hutchinson, C.R. and Floss, H.G.  
TITLE Direct Submission  
JOURNAL Submitted (30-DEC-1997) Dept. of Chemistry, University of  
Washington, Box 351700, Seattle, WA 98195-1700, USA  
REFERENCE 5 (bases 1 to 90445)  
AUTHORS August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R.,  
Hutchinson, C.R. and Floss, H.G.  
TITLE Direct Submission  
JOURNAL Submitted (19-JAN-2001) Dept. of Chemistry, University of  
Washington, Box 351700, Seattle, WA 98195-1700, USA  
REMARK Sequence update by submitter  
COMMENT On or before Jan 22, 2001 this sequence version replaced  
gi:2792340, gi:2832263.  
FEATURES  
source  
1. .90445  
/organism="Amycolatopsis mediterranei"  
/mol\_type="genomic DNA"  
/strain="S699"  
/db\_xref="taxon:33910"  
423. .1451  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAC01706.1"  
/db\_xref="GI:2792310"  
/translation="MAPTGGVGSRYVKTQLQLEVALIGAGTLARLHLEAWLGAGAAV  
RVYSDGRSRELAEECAKAGSLEALDGAADVICTPASHHEIATATAAGGVV  
CEXPDLASAEAEIIVTAERAGVRLYAHADVPAAPYARLHELIVASGRIGEGALGRF  
SFAYRPRWTHASAKSGGILTDQLLHDLADLAHVFGDVVRVHACVQGGIATPPEG  
AVATGATVLTHASGALSQVYSRWATPRPVRVAFVHVGSGSVSDSEWPQVRVVD  
GAGNFAYGGPSFVDEMRBFATAGGPEPRICAKDALAAVRIIHAAAEBSATGAV  
ELFVRGAA"

CDS  
1448. .2215  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAC01707.1"  
/db\_xref="GI:2792311"  
/translation="MKVALTSTPQAYAGALRGLPDVEVVAASWDAFRPVRAAAG  
ARVLCEYPPAAKETDLKAMIDRAGDELTFASPACGCEAEANVRKGIADGGIGELTIVL  
GSVAISDGVIGAAAPYLLDLADAVLGGEPAQQVYAOITNIVLSGRIGSEAVALTYR  
SQGVASFDCRRHGSATGLPAVTFIGDQGSVQDAGPGLGGERPELGGEDLEALMLKD  
FLGAGDGPDPGQAAALTRFIIQAAYESAHTGPVDL"  
2353. .2619  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAC01708.1"  
/db\_xref="GI:2792312"  
/translation="MQDRSVLGDPPKQVEAAMSVTQVKIALSGGPAELAPAHRSVDSG  
ALENTLKIRHGAGYEHFVHGGEFQAVDGCDAVFRWSHRTKIAE"  
2652. .3842  
/note="ORF0; similar to Streptomyces Sp. Chop"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450 monooxygenase"  
/protein\_id="AAC01709.1"  
/db\_xref="GI:2792313"  
/translation="MTDAISFEVPWDRDKFPPPAVDFSLRERFLAKMVYDPGHVGM  
LVTSYELVRELSDLRFHSCEVGHFPVTHQGVIPTHPLIPGMFIHMOPPEHTRRK  
ILTGFTVRRASRLIPRAEAAEQIEVNRKAGAPADVMDFAKPLVLRMLGELGLP  
YERDRYPAVTLHDAEADPAEAAAYEAGKFFDEVIERRRQDDLLISLVTED  
LTQELERNIVTLTFAGYETTGALATGVFALLHHTDQLAARAEPEKLDAAIEELLR  
VYVNGQHYRTALDEDVKLEGLIKGDVTVYSLPAANDPDAKFGCEALDIERTDSG  
HVAFGGHQCLQGNLARIELRAGFTALLRAPSELRALVPADEVPRLKGSVFSVKL  
PVSWM"  
gene  
3883. .18090  
/gene="rifa"  
CDS  
3883. .18090  
/gene="rifa"  
/note="rifa"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase"  
/protein\_id="AAC01710.1"  
/db\_xref="GI:2792314"  
/translation="WRTDLIKPLHVALLENATRFACKPAFADHRTVTYGDLEARTR  
LAGHLAGLVGRHGRVAICLNKRSTVSEYFALLRAGAVGPLNPGSAETAELEHPTD  
SGATVVTDAQAARILAPHELVLLVTGDDVPEGAHSDYDELALSEPAADDLELDE  
PAMVFTSGTTRGPKGVSTQRCNLSVASCTVPFGLSDQDRVLPFLFRHSLSHIA  
CVLSATVGVASVRIADGSSADVMRLIEAESSTFLAGVPTTYHLLVRAARQRFSAFS  
LAIGLAGAVLGALESFEETFGVPLIDAYGSTETCGAITWNPDPARVEGSCGLAV  
PGVDVRVDPDGLDVPAGEGEVWVGPNMLGHNSEATAAAMRDGFWRTGLAR  
RDAGFTTCGRKELIIRGGANIHPGEVEALRVTDGVADAAGVGPVPHDTLGEVPA  
YVIGPTGFDPAALTEKCEQLSAYKVPDRILEVAHPRITASKIRGLLTDEPAQR  
FRDLGFTSLAIVELNRNLTEHTGLMLPASAVDHPTPAALAAVRAELIGITOAFAEP  
VVAADPEPIATVGMACPLPGGVASPEDLWELVAERVDVASEFPFGDRGDLSDLDPD  
RERAGTSYVGCGFLHDAGEFDAGFPGLSPREAVAMDPOORLILLESWALENAGVDP  
IALKIDTGVFSLGMGGQGSAGAPFELGFTVTGVASVAGSRVSVLIGLEGPATVV  
DTACSSSLVAMHLAALQRCESMALAGGVVTMATPFGSFVRQRLARCKAF  
AAAADGTGWSGEGVVVLERLSVARERHRIILAVLGSVAVNQDQSGANGTAPNGLSQ  
RIVRALAAGLAPSDVDVEAHGTGTTLGDPLEAQAALLATYGOERKQPLWLSKPSN  
IGHQAAGAGVGVIKWQALRHETLPPTLVKPTLEVDMWSAGAILLFEARWPNRG  
RPRAGVSFGVSGTNAHLILREAPAEVPAAPELVPVILVVSARSTESLSQAERLA  
SLLEGDVSUTEVAGALVSRAVIDERAVVAGSREAVTGLRALTAGSGTGGKVVWV  
FFQGTQWAGMGRLELIVSFAERIAECAALAPWIDMSLVLDGEDGLRDVLIQ  
PACFAVGLAAAVWESVGRPDVAVVGHSGEITAAACVSGALSLEDAKVALRSQAIA  
AELSGRGGSAAVAGREDVVRSLVDGVVRVAAVNGSPSVIAGDAHALDATLELSGEG  
IRVVRVAVDYASHTRVEDIRITLAETLAGISQAQAPVFFYSTVTSWTRDAGVLGG  
YTRNLRNQRFRGAATALLGQHTVFEVSAHPVTVPQLSELDTGALTIRREDGGL  
RLLASMLFVRGIDVDWTMAPAGWDLPTIYAFEHRYMLLEPASAGDPLIGT  
VYSTGSDRLTAVQWSRRAQWMDGLVFNALVEAATRLGAGTLPVVGELVADP  
VVLPRGSEVOLIVGCEGQERRRPIEVFSREADEPWRHAGTLPAAAAVPEPAA  
GDATDVTVAGLRADRYGTHPALLDAAVRTVVGDDLLPSVMTGVSLLASGATVTFP











VERGDARELADITAGBODLPLDLSLLPALTSWRRGNOKHLLDTRLYRVEVTRLSKP  
TAPVLGDTWLLVASDADADOPALDGLADALGSHGVRRLRLDSDSCADRAVLAEAL  
ARTADVADATVLSVLPDIDERDADDCPELTRGLALITVALVQALATGTAGRLWATRG  
AVSTNPADVPHPVQAAGWGLRGVAGLEPFWGLVLDLPQVDFDERAGORLGLILAVK  
DAPDGEDVALATGVSGRVLRVHTVEALPTAAEFATGTATGLITCGTGLGAEVAVWL  
ARAGAHVLVTSRRGPAPGAELAELEGGVSVVACDVADRDLAALVLTALPEE  
LPLTGVTAGVGHYPLDTLSTAEFAGLTAALAGAAHLDALARELDLDFVLFPSI  
ACVMSGNSQASCANAVLDALHRRARGLAATSVAMGPWAEAGMAADAVSETLRR  
QGLGLDPAPEMTELRRAVRODVTTVADVDMQRYAPLFTSARPSALAGIPEVRL  
AADETEQDGAASEVTVRVALAPAEQURLLDLVRTESATVGLHSSADAVPEGRAP  
RVGDFSLTRALRGLAATGLSLPSTMVDFYPTPLLELAQYLRAELIAGVLEAGVP  
ATGQADDEPITALIIGMACRPGVSSPQMDLVASGTDAISEFPVNRGMQTHGLDPP  
PRPGTGTQCGFTHEADPFTFGIPREALVMDPOQLLETFTWTFESFRAGTRP  
ETLRLTCTFTVGSSTVGLGAGDGTGHTMTGSSPSVLSGLSVTFGLGPAVTD  
TACSSSLVLAHLACQSLRNGESNLAVAGSATIMTTPNPTAFSORALAKDRCRKA  
DDAGMTLAEAGVGLVERLSDAQRNGHPVLVLRGSAINQDASNGLTAPNGPQQO  
VIRQALANARAPGDIDALEHGTGTPGLDPEAQLFATYGRDRDFESALLIGSVKS  
NIGHTQSAAGIASVIMNVALRSELPTLHADAPSHVDWAGTVRLLTQARAWPET  
GRPRAAVSFGISGTNAHVLEQAPVADPAERPAVAPVPIAGVFWPVVTVARSAA  
ALRGQAEFLIAHAETVGTALPAAGPLDIGLSVSARARFHRVAVVPPAGTDLAALR  
AVATDGPSPVARGVADVEGRTVFPPGQSQWVGMSQLLDESVAFERIAECHAAL  
AVATDMSLVLRGVVGAFLSERVDVVPQSPFAVMVSLAALMRSGVLPDPAVGHSSQ  
ETAAVVSAGLSRQAGVALRVALSQAIGRALAGRMMSVALSVLDLEPLVEFEGRV  
SVAANVSPRSVVGPEALDHALHARLTADDIARRIADVYASHQVEDLHEEFLV  
LAELAPRTSEVFEFTVTDGMDLTARMAGYFNRNLGRVRFADAVADLILAEYRAFV  
EVSHPVLMAVQFEATDEAGVPAVAAAGTLRRDQGTDRLLSAAEVFVRGVVDWAGL  
FRGTGASRIDLPTFYAQHLEHMAVPAPEVAAADPDDAAFWTAVEDGDSVLTALAG  
TIEDSVAAVLPALTSWRARRDRSVDAWRYRVANKPLGLTLPHPSLTGTWLLYTAGD  
IDDTVAGALETYGAERVLRLDDEECVDRVLRERLAGAEDVTGIVSVLAAARBTDAV  
PGLTSLVLTALTVALTQALGDABEDIPAVFWALTGAVSTGRADELTPAQVQTGIGMT

Query Match 66.3%; Score 332; DB 1; Length 123580;  
Best Local Similarity 79.0%; Pred. No. 3.2e-27;  
Matches 395; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 1 TCCTCTTCGTGTCGGCGTGCACATGGCGGTGCAGCCCTCGACGCGCGCGAGTGTCTC 60  
DB 35023 TCCTCGTGCCTGCTGACCTGCGCGTGCAGCGCTGCGCAACGCGAGTGTCTCG 120

QY 61 ATGGCGTCCGCGCGCGCGTGTGATGGGCACGGTTCGAGAGTTCGTGCGAGTTCCTG 120  
DB 35083 ATGGCGTCCGCGCGCGCGTCAACGTGATGGCCACACCGCGGCTCGCTGTGTCGAGTTCAGC 35142

QY 121 CGGCAGCGCGCGGTGGCCCGCGCGCTGCAAGCGCTTCGCGGACGCGCGGACGCG 180  
DB 35143 CGCAGCGCGCGTGGCCCGCGCGCTGCAAGCGCTTCGCGGACGCGCGGACGCG 35202

QY 181 ACCGGCTGGTCCGAGGCGGTGGCGTCTCTGTTGAGCGGCTGTCCGAGGCTCAGCGT 240  
DB 35203 ACCGGCTGGTCCGAGGCGCGCGGTGCTGTGTCGAAACGGCTCTCCGACGCCACGCG 35262

QY 241 CGCGGGACACAGTCTCGCGGTGGTTCGGGGTTCGGCGGTCACTCCGAGCGGCGCTCG 300  
DB 35263 AACGGCCACCGGTGCTCGCGGTGCTCGCGGTTCGCGGTTCGCAACCGGACGCGCGCTCC 35322

QY 301 AACGGCTTGAACGCGCGCTCCGAGCGCGGTGATCCGCAAGCGCACTGCGCC 360  
DB 35323 AACGGCTTCAACGCGCGCTCCGAGCGCGGTTCGCGGTTCGCGGAGCGCGCTCGCC 35382

QY 361 GCCCGCGGACTGTCCACATCGGACGTTCGACGCGGTGGAGGCGGACCGGACCGGACG 420  
DB 35383 AACCGCGGCTGGCCACCGCGGACATCGACGCGGTTCGAGGCGGACCGGACCGGACCG 35442

QY 421 CTGGGCGACCGGATCGAGGCGCGGTGCTGGCCACCTACGGCCAGAAACGGGAAACG 480  
DB 35443 CTGGGCGACCGGATCGAGGCGGACGATCTCTGCGCCACTACGGCCAGGACCGGCGCCAC 35502

QY 481 CCGCTGTGGCTCGGTTCGGT 500  
DB 35503 CCGGTGTGCTCGGTTCGAT 35522

RESULT 7

AX211739 125401 bp DNA linear PAT 06-SEP-2001  
Sequence 35 from Patent WO0159126.  
AX211739  
AX211739.1 GI:15523950  
Streptomyces noursei  
Streptomyces noursei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomycetes.  
1  
Zotchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, F.,  
Stroem, A.R., Walla, S., Ellingsen, T.E., Sletta, H.V. and  
Gulliksen, O.M.  
Gene cluster encoding a nystatin polyketide synthase and its  
manipulation and utility  
Patent: WO 0159126-A 35 16-AUG-2001;  
Norges Teknisk Naturvitenskapelige Universitet (NO); STIPTILSEN  
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO);  
ALPHARMA AS (NO); Sinvent AS (NO); Zotchev, Sergey Borisovich  
(NO); Sekurova, Olga Nikolayivna (NO); Fjaervik, Espen (NO);  
Brautaset, Trygve (NO); Strom, Arne Reidar (NO); Walla, Svein  
(NO)  
Location/Qualifiers  
1. .125401  
/organism="Streptomyces noursei"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1971"  
/note="ATCC 11455"

Query Match 66.3%; Score 332; DB 6; Length 125401;  
Best Local Similarity 79.0%; Pred. No. 3.2e-27;  
Matches 395; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 1 TCCTCTTCGTGTCGGCTGCACATGGCGGTGCAGCCCTCGACGCGCGAGTGTCTC 60  
DB 35431 TCCTCGTGCCTGCTGCCCTTCACCTGGCGGTTCAGCGGTTCGCGCAACGCGAGTGTGCGC 35490

QY 61 ATGGCGTCCGCGCGCGGTGTGATGGGCAACGTTCGAGAGTTCGTGCGAGTTCCTG 120  
DB 35491 ATGGCGTCCGCGCGCGGTCAACGTGATGGCCACACCGGGCTCGCTGTGTCGAGTTCAGC 35550

QY 121 CGGCAGCGCGCGGTGGCCCGCGCGCTGCAAGCGCTTCGCGGACGCGCGGACGCG 180  
DB 35551 CGCCAGCGCGGTGGCCCGCGGACGCGCGGTTCGCGGACGCGCGGACGCG 35610

QY 181 ACCGGCTGGTCCGAGGCGGTTCGGGTGCTCTCTGTTGAGCGGCTGTCCGAGGCTCAGCGT 240  
DB 35611 ACCGGCTGGTCCGAGGCGCGCGGTGCTGTCTGTGAAACGGCTCTCCGACGCCACGCG 35670

QY 241 CGCGGGACACAGTCTTCGCGGTGGTTCGGCGGTTCGCGGTTCGCGGAGCGCGCTCG 300  
DB 35671 AACGGCCACCGGTGCTGCGCGGTGGTTCGGCGGTTCGCGGTTCGCGGAGCGCGCTCC 35730

QY 301 AACGGCTTGAACGCGCGCTCCGAGCGCGGTTCGAGCGCGGTGATCCGCAAGCGCACTGCGCC 360  
DB 35731 AACGGCTTCAACGCGCGCTCCGAGCGCGGTTCGAGCGCGGTTCGCGGAGCGCGCTCGCC 35790

QY 361 GCCCGCGGACTGTCCACATCGGACGTTCGACGCGGTGGAGGCGGACCGGACCGGACG 420  
DB 35791 AACCGCGGCTGGCCACCGCGGACATCGACGCGGTTCGAGGCGGACCGGACCGGACCG 35850

QY 421 CTGGGCGACCGGATCGAGGCGCGGTGCTGGCCACCTACGGCCAGAAACGGGAAACG 480  
DB 35851 CTGGGCGACCGGATCGAGGCGGACGATCTCTGCGCCACTACGGCCAGGACCGGCGCCAC 35910

QY 481 CCGCTGTGGCTCGGTTCGGT 500  
DB 35911 CCGGTGTGCTCGGTTCGAT 35930

RESULT 8









```

AVIGAEVDTDLRRRTATCDASVNAIGLESIAIRAGLLGDEPHQGLHFHAFVLLQTLFD
ELPRDRRLHAKADALVARRKHVGDEIEIRVAHSHWAKSPAAAEVLPRLLIAL
EBAEOAMAEYOEWIWLRAVHLVGLFPDDPATVLEOKLHLOLGOVLATRGYGHPE
AETATRGALSTVTHAPDPVIAWALCAASIVTGRYDASRQLSGLRDLSSQWTRPVP
AVIGRAIGEVIVHVGMPQALAELEHGERADHPANEGHALARTFQHDPRVCKRSY
NTFTHWLLGRIADARRELLSLYEYSPRSDRFALYVDGVVAWEAGDTRTLASS
DEGVAVGHGILLYKMLGVVKGWGLVHAGDRDGLALMHTSLAELGQSRTYLRHPL
HLGLLOAQOHNGQIEEARTTFRMLAAVEQRERVYLLPALPATPLLHLLHGHADE
ALADRDPRGQV"
/2190..3566
/gene="pink"
/2190..3566
/gene="pink"
/function="putative glucosyl transferase"
/codon_start=1
/transl_table=11
/product="PimK protein"
/protein_id="CAC20918.1"
/db_xref="GI:12055069"
/db_xref="GOA:Q9EWA4"
/db_xref="SPTREMBL:Q9EWA4"
/translation="MESARRPILLFVSLPESGILLNLLVLAGELSRQGVEDLWFATDEP
RENDVKRIAEGSPFASIGEVDSMSATVNSDEIVREVTQPSRKAHRAVVRHYTP
GLQAKFRRLQAVIDEVQALMWIDICISFAVDAALARNIPVLSVFFLPSPVTHYRT
HFAKSYPRGFVPVHTGLSRMTLAQRVANELLFKLRTFAMFNPLRGLKVLABDNRRN
ELGLPKAFMARIEHADLVLCNSLAELDFPDIPEKMRIVGAMVPPLEAPDQDLRS
WLDQASVVYVGLGTITRLTRREQVGSVMEVARRLDHQLVKLPSEQOHLPPRRSL
PQNLAVSVPMQMDVLAPHVKVFTFHGGNGENEGMYFGKPLVRLPWLVPDXYQAV
RQDQGLSLTLDRPQTIDVNDVVKLTVLVTFSPEYKAEERAAALMRSAGGRETAGL
VLSLPALA"
complement(3662..9736)
/gene="pims4"
complement(3662..9736)
/gene="pims4"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="PimS4 protein"
/protein_id="CAC20919.1"
/db_xref="GI:12055070"
/db_xref="GOA:Q9EWA3"
/db_xref="SPTREMBL:Q9EWA3"
/translation="MPNDKNDKLVLYLKWNTDLHQTRQLREVESAKQBPPIAIVG
MSCRLFGVRSPEEFWQLKDRDGIPTFTDRGMDLDLAGDGSSTAAEGFDIDA
ASFDAFGFISPREAVTMDPQORILLSEWEALERAGIDPPLRGTEGVFVGTSGID
YINTVNSREDITEGHTGLIASILLGRVSYTFGLEGPATVTDTCSSLSVLSHLAAQ
ALRVGCSLALAGGVTAISTPMFSVYGFSGGLAPDGRCKAFSDTADGTAWSEGVVL
VVERLSDARRNGHPVLAVRGSVAVQDGSNGITAPSGSQQRVIRQALAAAGLSASD
VDVZAHGTALGDFIEHQALLATYGRDRDPERPULLGSVKSNGHTQAAAGAAGLI
KMLAMRHGVLPTLHVEAPSHVMSGALQLLTPQWPKPADRPWRAGVSFAISG
TNAHILIEQPPAAEPDEBEGPHTSPSAVPLVLSAKSEALTAQLDRISALTDRSPLO
LGLSLATTRAHLEHRAVLSEGGETVEVARHAEHTAGKLAVLFSQGAORVGMGREL
YGRFPVFAELDGLVLAHPDAGRLDVMFGEAEGLDGTGTQPALFALEVALFPLSLSIG
YCRFPVAGHSIGELAAHVAGVFSIADACALVAARLMOALPAGAMVAQAAESEV
AARLVDSVIAAGAAHVAVIAGEEREVRIAEFTGLGRKTRRLFPVSHAFSPIMEP
MLEDFFRAVLEGSYEPRIALVNSLTGELATVPALKRDSRSEETALITALAQLHVAGV
AAEGASVFLGPDGVLTAMAGTLDGTATVPALRKSDEETALITALAQLHVAGV
VWMSGIGAGTGAARVDLPYTFQHERYWRPAPVLSGVVVGAGLRAEHLPLIGATILTA
BSGEVILTGRLSLOAVPWLTDHMGSSVLPFATPLELAMRAADQVGDVDEFLVMT
PLVLAESCTQVVLGPADEAGTRTISHARPDPDQPMQHQAGGTLISGERIVDF
GSTVPMQAEAVLBGFDTTIGYFAPGLRSWQCEGAYVEVALPQVAEDAGAF
GLHPVLLDVAQSRMAGIAGADGTFLPFAMRGSVSLHAGGASVLRARVTKTGEDTSL
VADAGELGVASVATLADLSDGDFVVPVRSVGDDVPAAAHELTARTLGLLOEWSG
VLSIADGVAGSVAATLADLSDGDFVVPVRSVGDDVPAAAHELTARTLGLLOEWSG
REGTSARLVFYTRNVSADGGSVEDVAAAVMGLVRSQAQSPGRIALVLEADSG
DLAEVLGLPGLLAAGDAQFVVRBGAVRGRVLSNEGAGVSPREWDAGTIVLTGTT
GLGLRELARHLVVERGRVLLVSRGPEASGELRDELAAGABEIVRACDVADRA
AYEAVGVADAHPHTAVTHTAGVLDGVIASILTPELRSANVLRPKVDAANWHEATKO
LDLAEVYSSTISGVLGSGQNYAGNVFLDALAGHRAGLGLPAQSLAWGAWTQGG
MTATLSDADQMORWASGTAPLSPERGLAFDAATAGSATHAVAGVSGVPRAGQVPP
SMRLGLVKARERTAVNTTGETAVSLADRLRAARPEERTQMLTDVNRVTEAAVIGHSS
QNVDAAREYELGFDLSLTVELNRNLSVTIGRLSATVIGLVDVFAAETGRTFPAALDQTYG
QSALDGSASQAGFLPVAAPETDLSLRFMDALDLSGKVPQAQRMLSALGALRFPSPNTA

```

```

ELEDLPLPATLAEGPGTPELICI VSTPVGGVHVEVARLAASFRGERHSHVPLVGFAT
GERLPATATVAVTAEASALRASDGNPFWLVGHSSAGAFAYLAALLENWGTPEAPV
VLLDTLSLRHEONESIDYAGLMRRHFMWDEVPFMTNLSRLSAMARWGMMLQLEVRH
TTPVLLIIRATKETFTGIDTDTIGYEDHGHSPVDVRSVDADHFSWVRDDAPETARI VRE
WLDLSGNA"
complement(9772..15198)
/gene="pims3"
complement(9772..15198)
/gene="pims3"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="PimS3 protein"
/protein_id="CAC20920.1"
/db_xref="GI:12055071"
/db_xref="GOA:Q9EWA2"
/db_xref="SPTREMBL:Q9EWA2"
/translation="MTTSETVVAALRASLKEAENLRQNRKLVAATEPIAVAMAC
RYPGVQSPEDLWELLTAGKDAIGAFPTDRGMDLDALRDAGVDNRGNVQSQGGFLDG
VGDVADFPGISPRESVTMDPQORILLLETSWEAIERAGIDPTTLRGSRGTGFTNGO
DYAVLLRSLADATGDI GTIAASATSGRLSVTLGLEGPATVDTACSSLSVALHWAS
HALRAGCSLALVGVVWMSFSGSLMERSQGLLAADRECKAFSDTADGTGNSGVGV
LVVERLSDARRNGHPVLAVRGSVAVQDGSNGFTAPSGSQQRVIRQALAAAGLSAS
DVVVZAHGTGTPLGDPTEAQALLATYQDRDPERPULLGSVKSNGHTQAAAGAAGL
IKMILAMRHGVLPTLYADVPSSHVDVTSWEVLELITEPOEWPKADRPWRAGVSFSGIS
GTNAHVIVQAPDEPEVPAEGLPAVVPVLSAKTEERAAQLARISRRRTGTSFVD
VGFSLASGRSAPEHRAVLLAQCPCDGLSEAGRAETSRPLAVLFSQGAQVGMGRE
LYVRFVPAEALDEVLAFHFDAGLRDVMFGEAEGLDGTGTQPALFALEVALFRLAESL
GYRPFVGHSTIGEIAAHVAGVFSITDICALVAARLMOALPAGAMVAQAAESE
VAARLVDSVIAAVNGPEAVVLGADDETAVLRAEELAAAGRKTORLPVSHAFSPIME
PMLEDFFRAVLEGSYAGQIPLVNSVTGELAVEELVCSDPYVWRVRETRFADGVSA
LEAGASAFLELGPDGVLTAMARHTLDETKVAPVLRDORPEETALLTAHLHVAGI
GVDSGIGIPAGTGARVDLPYTFQHERYWRPAPVLSGVVVGAGLRAEHLPLIGATIL
AESGEVLTGRLSLQTHPLTDHAI GSNVFFPGTGTGLBLAMHAAQAGCDRVDLVM
TFLVLAENGTOQVVLGPADEAGTRTISHARPDPDQPMQHQAGGTLISGERVVD
PGLVWPPQDAVAVDLEFEGTGYGPAFLSRVSVMSGCGAYAVLPHDVAEDVADR
FGLHPALLDAVLQSHRIAGVSGADQFLPFAWRGVS LHAGGASVLRVTVKTCDDSMS
TAAVDAEGAPVLSAEAMVARYDRPTGQAAARHGAERDELLSLEWMSGPEVPPAGVR
CVSLGADLGVGASVASIADLSDGDFVVLVPGVGVDPVPAAAHELTARTLGLQEWS
GREGSGRLVPTVTRNVSADGGSVEDVAAAVMGLVRSQAQSPGRIALVLEADSG
GDLEVLGTLGGLLAAGDAQFVVRBGAVRGRVLSNEGAGVSPREWDAGTIVLTG
TGLGLRELARHLVVERGRVLLVSRGPEAFVSGLSDELAAHAEVTVRACDVADR
AAVHALVAAPAEHPLTAVITAGVLDGVIASILTPELRSANVLRPKVDAANWHEATK
DLDLAEVLYSSISIGVLPAGQANYAGNVFLDALAWRMGQGLPAQSLAWGAWTQGG
GMITATSDTMORWASGTAPLSPERGLAFDAACARAAHLVAVRGVGVPRAQVAPV
PSVLGLVKGARERTAVNTTGETAVSLADRLRAARPEERTQMLTDVNRVTEAAVIGHAS
AKVNDAREFNELGFDLSLTVELNRNLSVTIGRLSAAVLFYDPNPTALADHFVSELV
BEAATGPALLAELERLDSALAGADPAPTRAATRAAQILDLDWDHDAFPAEDGAEVAE
RLEAASDTDEITFAFIDNELGRHSDR"
complement(15210..43733)
/gene="pims2"
complement(15210..43733)
/gene="pims2"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="PimS2 protein"
/protein_id="CAC20921.1"
/db_xref="GI:12055072"
/db_xref="GOA:Q9EWA1"
/db_xref="SPTREMBL:Q9EWA1"
/translation="MDQOKRLDYLKRASADLRRSQRQVGELEAAAREPPIATVGMSCR
FPGVRSPEELWSLILAADGDI TKSADRGWELLAEADAGEFSGFLHDAABEPDASFFG
ISPREALMDPQORILLSEWEAEPFAGIDVPAVGTGTFMFGVAMPQEPVGRPDDVY
QGFALTGTTTIVISGRLAYFFGAVGPAVDTACSSLSVALHLSRQGCCSIALA
AGTVWMSGIGTTFVFNQGGLSADGRCSFADSGTGWSEGVGLVLERUSEARNR
HEILAVIRGSVAVQDGSNGITAPSGSQQRVIRQALVRSADDEVTEVAHGTTGT
LGDVPAQALLATYQDRDPERPULLGSVKSNGHTQAAAGMGLVKMWLAHRHGVLP
RTLHVDAQSSHVDVTDGQAVRLLTTHVPMWQCEQPRAGVSSGISTNAHTIILEAPA
RPSFRAGQEDTAPVAPLAAAGVPMVLSGRTEALRAQAARLLEHLTSRPGLAFLDVA
YSLATRSGLEHRAAFTAAAGPDGARAALTALADGTPAGLVDOTARTGSKLAFLFAGQ
YSQFPMGRMLEAARFPFATALDEVLHFDGLERPLKDVLPFAAETGRTFPAALDQTYG
AQPALFALEVALYRLAESFGITPDPLAGHSIGETIAAAHVAGVFTLADAAALVLRGLR

```



MOALPEGAMVSLSEATEBEVLLLEDRPRLSIAVNGPAAVIAAGEADDDVATAVEHF  
DALGRRTRKLVSHAFSLPMPEMLAEPTAVSRLTPOAPVIVPVVGLTGLTATVEOL  
TSPOYLDHARHAFADALARRRHLPRMGPDGLSAMQSCCRGHGRRRR  
LRHARRVVRPEAETLTAALAGLHTRGVAWMEPYFGTGARADLITYAFQRRYY  
PKSLPATGGDVRAAGLGNHPLITAASVANSGLLTLGRSLRTHFWLADHVRGT  
VLLPPTAFLEAVRAGDAGCGRVEELTLAAPLVLPERGGVQVQVWVGSPDASGRRV  
SVHRPDGPELPWTQHAAGTLVAGEHHAGFDATVWPPADATPLDLOGFYARMADTGF  
GYGLFQGLRAARMSGDVYAEVALPDSIDGAESTESSAAGLHPALLIDLAALHAAAFVD  
LGEDARGGLPFSWQDVTLHAGSASTVRVKLTDPDGDVALAVADTTGAPVASIGLVL

Query Match 65.5%; Score 328.2; DB 1; Length 84985;  
Best Local Similarity 79.8%; Pred. No. 9.1e-27;  
Matches 387; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTGGTGGCTGCACATGCGCGTGCAGGCGCTGCAGCGCGGAGTGTCTCC 60  
Db 75953 TCCTCCTCGTGGTGGCTGCACATGCGCGCGTGCAGGCGCTGCAGCGCGGAGTGTCTCC 75894

QY 61 ATGGCGTTCGCGGCGGCGTGTATGATGGCAAGTGCAGAGCTTCGTGAGTTCCTCG 120  
Db 75893 CTGGCGGTTCGCGGCGGCGTGTATGATGGCAAGTGCAGAGCTTCGTGAGTTCCTCG 75834

QY 121 CGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180  
Db 75833 CGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 75774

QY 181 ACCGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
Db 75773 ACCGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 75714

QY 241 CGCGGCGACCAAGTCTCTCGCGTGGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
Db 75713 AACGCGCACCAAGTCTCTCGCGTGGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 75654

QY 301 AACGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
Db 75653 AACGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 75594

QY 361 GCCCGCGGCTGTCCACATCGGAGCTGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
Db 75593 AACCGCGGCTGTCCACATCGGAGCTGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 75534

QY 421 CTGGCGACCGGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
Db 75533 CTCGCGACCGGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 75474

QY 481 CCGCT 485  
Db 75473 CTGCT 75469

RESULT 11  
AB070940 104326 bp DNA linear BCT 22-JAN-2002  
DEFINITION Streptomyces avermitilis oligomycin biosynthetic gene cluster.  
ACCESSION AB070940  
VERSION AB070940.1 GI:15823967  
KEYWORDS  
SOURCE Streptomyces avermitilis  
ORGANISM Streptomyces avermitilis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
AUTHORS Omura, S., Ikeda, H., Iehikawa, J., Hanamoto, A., Takahashi, C.,  
Shinoe, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,  
Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M.  
TITLE Genome sequence of an industrial microorganism Streptomyces  
avermitilis: deducing the ability of producing secondary  
metabolites  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
MEDLINE 21477403  
PUBMED 11572948

REFERENCE 2 (bases 1 to 104326)

Ikeda, H.

Direct Submission

Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of  
Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1  
Shirokane, Minato-ku, Tokyo 108-8641, Japan  
(E-mail: ikedamc.pharm.kitasato-u.ac.jp, Tel.: +81-3-5791-6242,  
Fax: +81-3-3444-6197)

Location/Qualifiers

1. 104326

/organism="Streptomyces avermitilis"

/mol\_type="genomic DNA"

/db\_xref="taxon:33903"

complement(1..1134)

/gene="mcmA"

complement(1..1134)

/gene="mcmA"

/note="methylmalonyl-CoA mutase"

/codon\_start=1

/transl\_table=1

/protein\_id="BAB69185.1"

/db\_xref="GI:15823968"

/translation="MTGISTAADANARPRQLVGSQTRAVPVVFDLPTFAGLSDAPAA  
RGLVGRGVAVDSIDDMRVLFGLRLDRVPSLAVDAPAPAEFLVLYQVVAEHGAPAQ  
RLAGVRGVDFKEFLRPGVFPFLRPSLRILAMNYFVCAQELRWQCLSVCGHLAGT  
GADPALEVAFVAGGIEYLRAAVTAGLDVTVAPRLTCLVGGTGTGRNKLGTATRR  
VMTTANERFGSGRTAPRLYVTRPDGLATLACTIEAGATLVDRTERLGGTJTVL  
EHGLGOLLGPHGCDRPAATPRPHQDSLDLQARQTERLAKLAWRVQPAVDALLR  
LRXAAQDDNVLYPMREALAAATVGEVCDLAURSTWGTGSPSAT"

complement(1131..1886)

/gene="pptA"

complement(1131..1886)

/gene="pptA"

/codon\_start=1

/transl\_table=1

/product="phosphopantetheinyl transferase"

/protein\_id="BAB69186.1"

/db\_xref="GI:15823969"

/translation="MRAHEVADVPVSSLHPREALVQDAVESRRERFATGRHCARRA  
LASLGVPPGILLATPGCAPKPNAGVCCAIHCAGYRAAANAVARAGIARVGDADPRP  
LIDLVDAVALPAEPQVQKENTARHPGVVWVKLLFSAKESVIKAWFATGPGPLGFEA  
LLDLPVPSVTRVRLPTALRAAEPDLRHLGRLVIRHGIIVTAVVVDGPRDTPHAP  
SPYDRKHSAPTGRSADRPGRPTDRATRPTA"

complement(1910..2875)

/gene="simX4 homolog"

complement(1910..2875)

/gene="simX4 homolog"

/codon\_start=1

/transl\_table=1

/product="SimX4 homolog"

/protein\_id="BAB69187.1"

/db\_xref="GI:15823970"

/translation="MSQDRHDGSTAALGGPSHGRLLAVIDIHRHQRNDIIVDLHP  
GCADFMLIVAGDVASVADITWALSRLDRFAKVIWAPGNHFWLTPGDPDELGERK  
YRHLVDICRALGVVTPEDPYVWVGAGPVAVPLFLLYDYFTLPPGASKSHALRLA  
DDAGICTDFFLLHPDPYPSREACRARVRATRLDALPDLPTVLVNHPPVLPREPT  
RVLRRPEALWCGTTATADWHLRYRAAVVGHGHLIERTIECDVDPHEVSLGYPREW  
ORYGRPDWVYQILTPAPAPAPVPVPVAPPGHRTASADRRP"

3213..3971

/gene="olmC"

3213..3971

/gene="olmC"

/codon\_start=1

/transl\_table=1

/product="thioesterase"

/protein\_id="BAB69188.1"

/db\_xref="GI:15823971"

/translation="MSANRVGNLWYRQHPAPTATKLYCLPHAGGSASFVLPVSKA  
LAPGVDAIQVPRQDRHEKCTDITAEADAVTAVLELADRPVLVFGHSGATLTA  
FEVALRERKIVPLAFASGRAPSRHRDSEVHLRGDDETIAELKLTSGTRSEVIGD  
EELIRWVLPAIRSDYKAAETRYAGGARLTAP IAHAFGVEDPRVSLDEAQAWADHTTG  
HFELRSHPGGHFYINDQAPRVIEIARVSSVAS"

4643..7438

/gene="olmR1"



















Result No.	Score	Query Match %	Length	DB	ID	Description
1	501	100.0	53789	2	AAV211187	AAV211187 Amycolato
2	334.6	66.8	5676	2	AAV211186	AAV211186 Amycolato
3	332	66.3	125401	4	AAV217186	AAV217186 Streptomycin
4	328.2	65.5	20394	4	AAF24892	AAF24892 Pimaricin
5	325	64.9	9975	7	AAAL61173	AAAL61173 Actinosyn
6	325	64.9	82746	7	AAAL611224	AAAL611224 Actinosyn
7	320.8	64.0	113193	7	AAAD54645	AAAD54645 Streptomycin
8	320.2	63.9	9222	7	AAAL61171	AAAL61171 Actinosyn
9	320	63.9	14055	7	AAAL61170	AAAL61170 Actinosyn
10	317.6	63.4	65140	4	AAAL71184	AAAL71184 Streptomycin
11	314.4	62.8	77536	3	AAAL4651	AAAL4651 Nucleotid
12	307.4	61.4	15872	3	AAZ87283	AAZ87283 S. venez
13	307.4	61.4	43280	2	AAAT80413	AAAT80413 Ty lactone
14	307.4	61.4	103599	4	ABX04971	ABX04971 S. cinna
15	305.8	61.0	15872	2	AAAT68715	AAAT68715 Streptomycin
16	305.8	61.0	38506	3	AAAF75633	AAAF75633 Nucleotid
17	305.8	61.0	38506	3	AAZ56001	AAZ56001 Recombina
18	305.8	61.0	38506	7	ABBS6090	ABBS6090 S. venez
19	305.8	61.0	38506	7	ADA09418	ADA09418 Cosmid pK
20	304.2	60.7	4674	3	AAAL4666	AAAL4666 Nucleotid
21	304.2	60.7	4725	3	AAAL4665	AAAL4665 Nucleotid
22	304.2	60.7	4737	3	AAAL4668	AAAL4668 Nucleotid
23	304.2	60.7	4767	3	AAAL4667	AAAL4667 Nucleotid

4  
4  
4  
4  
4  
4  
4  
4  
4  
4

1000

18-AUG-1997; 97WO-EP004495.



```
Db 3897 TCCTCGTGGTGGTGGATGACACCTGGCGCGGAGCGCTGCGGACGGCGAGTGTCTCG 3956
QY 61 ATGGCGCTCGCGGCGCGCGTGTGATGGTGGGACACGGTGCAGACGTTCTGTCGATTCCTCG 120
Db 3957 ATGGCGCTCGCGGCGCGCGTGTGATGGTGGGACACGGTGCAGACGTTCTGTCGATTCCTCG 4016
QY 121 CGGACGCGGCGGCTGGCCCGGCGCGCGTGTGATGGTGGGACACGGTGCAGACGTTCTGTCG 180
Db 4017 AAGCAGCGGCGGCTGGCCCGGCGCGCGTGTGATGGTGGGACACGGTGCAGACGTTCTGTCG 4076
QY 181 ACCGGTGTTCGAGGCGGCTGGCGGCGCGTGTGATGGTGGGACACGGTGCAGACGTTCTGTCG 240
Db 4077 ACGGCTGTGGCGGCGGCGGCTGGCGGCGCGTGTGATGGTGGGACACGGTGCAGACGTTCTGTCG 4136
QY 241 CGGCGGACACGAGTCTCGCGGCGGCGGCTGGCGGCGCGTGTGATGGTGGGACACGGTGCAGAC 300
Db 4137 CGGCGGACACGAGTCTCGCGGCGGCGGCTGGCGGCGCGTGTGATGGTGGGACACGGTGCAGAC 4196
QY 301 AAGCGCTTGACGCGCGCGCGGCGGCGGCTGGCGGCGCGTGTGATGGTGGGACACGGTGCAGAC 360
Db 4197 AAGCGCTTGACGCGCGCGCGGCGGCGGCTGGCGGCGCGTGTGATGGTGGGACACGGTGCAGAC 4256
QY 361 GCGCGCGGACTGTTCACATCGGACGTCGACGCGGCTGGCGGCGCGTGTGATGGTGGGACACGG 420
Db 4257 GCGCGCGGACTGTTCACATCGGACGTCGACGCGGCTGGCGGCGCGTGTGATGGTGGGACACGG 4316
QY 421 CTGGCGGACCGGATCGAGGCGGCGGCGGCTGGCGGCGCGTGTGATGGTGGGACACGGTGCAGAC 474
Db 4317 CTGGCGGACCGGATCGAGGCGGCGGCGGCTGGCGGCGCGTGTGATGGTGGGACACGGTGCAGAC 4376
QY 475 GAAACGCGCTGTGGCTTCGCGGTCGCGTG 501
Db 4377 GAGACGCGCTGTGGCTTCGCGGTCGCGTG 4403
```

## RESULT 3

```
AAD17186
ID AAD17186 standard; DNA; 125401 BP.
XX
AC AAD17186;
XX
AC AAD17186;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; ds.
XX
OS Streptomyces noursei.
XX
FH Key Location/Qualifiers
CDS 6337..34771
FT /*tag= a
FT /product= "NysI complete protein"
FT 34792..51099
FT /*tag= b
FT /product= "NysJ protein"
FT 51155..57355
FT /*tag= c
FT /product= "NysK protein"
FT 57503..58687
FT /*tag= d
FT /product= "NysL protein"
FT complement(58786..58980)
FT /*tag= e
FT /product= "NysM protein"
FT /note= "CDS does not include start codon"
FT complement(59045..60241)
FT /*tag= f
FT /product= "NysN protein"
FT /note= "CDS does not include start codon"
FT complement(60238..61296)
FT /*tag= g
FT
```

```
FT CDS
FT /product= "NysD2 complete protein"
FT 120628..121308
FT /*tag= h
FT /product= "NysR4 (long) protein"
XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB0000509.
XX
XX 08-FEB-2000; 2000GB-00002840.
XX 10-APR-2000; 2000GB-00008786.
XX 14-APR-2000; 2000GB-00009387.
XX
XX (UYN0-) UNIV NORGES TEKNISK NATURVILTENSKAPELIGE.
XX (SNTF) SINTEF STIFTTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIELEWSKA H.
XX (ZOTC/) ZOTCHEV S B.
XX (SEKU/) SEKUROVA O N.
XX (FJAE/) FJAEVRIK E.
XX (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX (VALL/) VALLA S.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI; 2001-557614/62.
XX P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX AAE10149, AAE10150.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides, useful
XX as antibiotics and antifungals.
XX
XX Claim 1; Page 188-254; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrolide antibiotic nystatin. The
XX nystatin PKS is useful as antifungal antibiotics. The present sequence is
XX a Streptomyces noursei nystatin PKS gene cluster DNA
XX
XX Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 U; 0 Other;
XX
XX Query Match 66.3%; Score 332; DB 4; Length 125401;
XX Best Local Similarity 79.0%; Pred. No. 1.8e-43;
XX Matches 395; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
QY 1 TCCTCTTCGCTGTCGCGCTGCACATGCGGTGCAGGCGCTGCAGCGGCGAGTGTCTCC 60
Db 35431 TCCTCTTCGCTGTCGCGCTGCACATGCGGTGCAGGCGCTGCAGCGGCGAGTGTCTCC 35490
QY 61 ATGGCGCTTCGCGGCGGCGTGTGATGGTGGGACACGGTGCAGACGTTCTGTCGATTCCTCG 120
Db 35491 ATGGCGCTTCGCGGCGGCGTGTGATGGTGGGACACGGTGCAGACGTTCTGTCGATTCAGC 35550
QY 121 CGGACGCGGCGGCTGGCCCGGCGCGGCTGTGATGGTGGGACACGGTGCAGACGTTCTGTCG 180
Db 35551 CGGACGCGGCGGCTGGCCCGGCGCGGCTGTGATGGTGGGACACGGTGCAGACGTTCTGTCG 35610
QY 181 ACCGGTGTTCGAGGCGGCTGGCGGCGGCTGTGATGGTGGGACACGGTGCAGACGTTCTGTCG 240
Db 35611 ACCGGTGTTCGAGGCGGCTGGCGGCGGCTGTGATGGTGGGACACGGTGCAGACGTTCTGTCG 35670
QY 241 CGCGGACACGAGTCTCGCGGCGGCGGCTGGCGGCGGCTGTGATGGTGGGACACGGTGCAG 300
Db 35671 AAGCGCACACCGGCTGTGCGGCGGCTGGCGGCGGCTGTGATGGTGGGACACGGTGCAG 35730
QY 301 AACGGCTTGACGCGGCGGCGGCTGGCGGCGGCTGGCGGCGGCTGTGATGGTGGGACACGG 360
FT
```

Db 35731 AACGGCTTACCGCCGCCCAAGGCCCTCCAGCAGCGGCTCATCCGCCAGGCGCTCGCC 35790  
QY 361 GCCCGCGAGCTGTCCACATCGGACATCGACGCGGTGGAGGCGCACCGGCACCGGACGACC 420  
Db 35791 AACGCGCGCTGGCCACCGCGGACATCGACGCGGTTCGAGGCGCACCGGACCGGACCGCCG 35850  
QY 421 CTGGCGGACCGGATCGAGCGCGAGCGGTCTGGCCACCTACGGCCAGAACCGGGAACG 480  
Db 35851 CTCGGCGACCGCCATCGAGCGCGAGAGCATCTCCGCCACTACGGCCAGGACCGCGCCAC 35910  
QY 481 CGCGTGTGCTCGGCTCGGT 500  
Db 35911 CCGGTGCTGCTCGGCTCGAT 35930

## RESULT 4

AAF24892  
ID AAF24892 standard; DNA; 20394 BP.  
XX  
AC AAF24892;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Pimaricin biosynthesis is associated polyketide synthase gene.  
XX  
KW Polyketide synthase; oxidative modification; metabolite; antibiotic;  
XX anticancer; pimaricin; ss.  
OS Streptomyces natalensis.

Key Location/Qualifiers  
CDS I..20394  
FT /\*tag= a  
FT /product= "polyketide synthase"  
XX

WO200077222-A1.

21-DEC-2000.

14-JUN-2000; 2000WO-EP006227.

14-JUN-1999; 99EP-00201893.

(STAM ) DSM NV.

Martin JF, Aparicio JF, Colina AJ;

WPI; 2001-080693/09.

P-PSDB; AAB31558.

XX New polynucleotides encoding enzymes involved in the biosynthesis of  
PT pimaricin, useful for modifying the biosynthesis of pimaricin and in the  
PT synthesis of new compounds.

Disclosure; Page 53-80; 116pp; English.

XX The present sequence encodes a polyketide synthase which is associated  
CC with the biosynthesis of pimaricin. The polyketide synthase polypeptide  
CC is useful for the oxidative modification of a methyl group of a suitable  
CC compound, e.g. a bioactive compound including a secondary metabolite,  
CC antibiotics and anticancer agents. Recombinant cells comprising the gene  
CC are useful for the production of pimaricin. The polyketide synthase  
CC polynucleotide may be over expressed in Streptomyces, leading to an  
CC increase in the biosynthesis of pimaricin, as a source of primers for  
CC amplification reaction and as probes

Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 U; 0 Other;

Query Match 65.5%; Score 328.2; DB 4; Length 20394;

Best Local Similarity 79.8%; Pred. No. 9.7e-43;

Matches 387; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1 TCGTCTCGGTGCTCGCTGCACATCGCGGTGTCAGGCCCTGACGCGGCGAGTGTCC 60

Db 5200 TCCTCTCGTGTGCTGCTGCTGCATGATGCGCGCAGGCGCTGGCCAGGCGAATGCTCC 5259  
QY 61 ATGGCGCTCGCGGCGCGGTGATGTGATGGCAGCGTTCGAGACCTTCTGTCAGTTCG 120  
Db 5260 CTGGCGGTTCGCGGCGGTGTGACGGTGTGATGGCCACCCCGTCTCTCTGAGTTCGCG 5319  
QY 121 CGGACGCGCGGCTGGCCCGCCGACGCGCGCTGCAAGGCGTTCGCGCAGCGCGGACGCGC 180  
Db 5320 CGGACGCGCGGCTGGCCCGCCGACGCGCGCTGCAAGGCGTTCGCGCAGCGCGGCGC 5379  
QY 181 ACCGCTGTGCTCGAGGCGCTCGGGCTGTCTCTGTGTGAGCGGCTTCTCGAGGCTCAGG 240  
Db 5380 ACCGCTGTGAGCGAGGCGCTCGGCTGTCTCTGTGGAACGGCTCAGGACGCGCGCGA 5439  
QY 241 CGCGGCGACCGAGTCTCTCGCGGTGCTCGCGGCTGCGCGGTCAACTCGACGCGGCGGT 300  
Db 5440 AACGGCCACCGAGTGTCTCGCGGTGCTCGCGGCTTCCGCGGTCAACGACGCGCGCTCC 5499  
QY 301 AACGCTGTGACGCGCGCGAAGCGCGCTCCAGCAGCGCGTGTCTCGGAAGCACTGGCC 360  
Db 5500 AACGCTGTGACGCGCGCGAAGCGCGCTCCAGCAGCGCGTGTCTCGGAAGCACTGGCC 5559  
QY 361 CGCGCGGACTGTCCACATCGGACGCTCGACGCGGTGGAGCGCACCGGACCGGACGACC 420  
Db 5560 AACGCGCGGCTGGCGCGCTCCGAGGTTCGAGCGCGGTGGAGCGCACCGGACCGGACG 5619  
QY 421 CTGGCGGACCGGATCGAGCGCGAGCGGTCTGCGCGACCTACGGCGCAGAACCGGGAACG 480  
Db 5620 CTCGGGTGACCGGATCGAGCGCGAGCGGTCTGTCGCGCACCTACGGCGCAGGCGCGCTG 5679  
QY 481 CGGCT 485  
Db 5680 CTGCT 5684

## RESULT 5

AAL61173  
ID AAL61173 standard; DNA; 9975 BP.

XX AC

XX AAL61173;

XX 22-SEP-2003 (first entry)

XX Actinosynnema pretiosum polyketide synthase (PKS) gene #4.

XX Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;  
KW gene; ds.

XX Actinosynnema pretiosum.

XX WO2003045312-A2.

XX 05-JUN-2003.

XX 21-NOV-2002; 2002WO-US037547.

XX 21-NOV-2001; 2001US-0332158P.

XX (UNIW ) UNIV WASHINGTON.

XX Floss HG, Yu T, Leistner E;

XX WPI; 2003-493374/46.

XX Novel maytansinoid produced by bacterial host cell transformed with  
PT expression vector comprising open reading frame from ansamitocin gene  
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.

XX Disclosure; Page 62-67; 160pp; English.

XX The invention relates to maytansinoid produced by bacterial host cell  
CC transformed with expression vector comprising open reading frame from



CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
 CC useful as a potent antitumour agent. The present sequence is A. pretiosum  
 CC ansamitocin gene cluster I polyketide synthase (PKS) gene  
 XX  
 SQ Sequence 9975 BP; 947 A; 4102 C; 3795 G; 1131 T; 0 U; 0 Other;  
 Query Match 64.9%; Score 325; DB 7; Length 9975;  
 Best Local Similarity 78.0%; Pred. No. 3.4e-42;  
 Matches 391; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
 QY 1 TCGTCTTCGCTGCTGCGTGCACATGGCGGTGCAGGCCCTGCAGCGCGCGAGTGTCTCC 60  
 Db TCGTCTTCGCTGCTGCGTGCACATGGCGGTGCAGGCCCTGCAGCGCGCGAGTGTCTCG 5550  
 QY 61 ATGGCGCTCGCGGCGGTGATGATGGGACCGTGCAGACGTTCTCGAGTTCCTCG 120  
 Db ATGGCGCTCGCGGCGGTGATGATGGGACCGTGCAGACGTTCTCGAGTTCCTCGAC 5610  
 QY 121 CGGCAGCGCGGCTGGCCCGGCGGTGCAGGCGTTCGCGACGCGCGGCGGACGCG 180  
 Db CGGCAGCGCGGCTGGCCCGGCGGTGCAGGCGTTCGCGACGCGCGGCGGACGCGC 5670  
 QY 181 ACCGGCTTGGTCCGAGGCGGTGCGGCTGCTCTGTTGAGGCGGTGCTCCGAGGCTCAGCGT 240  
 Db ACCGGCTTGGTCCGAGGCGGTGCGGCTGCTCTGTTGAGGCGGTGCTCCGAGGCTCAGCGT 5671  
 QY 241 CGCGGCGACAGTCTTCGCGGTGCTCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCG 300  
 Db CGCGGCGACAGTCTTCGCGGTGCTCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCG 5730  
 QY 5731 AAGCGGACGCGCTGTGCGGCGGTGCTGCGCGGCGTCACTCCGACGCGCGGTCC 5790  
 QY 301 AAGCGGCTTGACGCGCGCGGCGGCGGCGTCCAGCGCGGTGATCCGCAAGGCACTGGCC 360  
 Db AAGCGGCTTGACGCGCGCGGCGGCGGCGTCCAGCGCGGTGATCCGCAAGGCACTGGCC 5791  
 QY 361 GCGCGCGGCTGCTCCATCGGACGTCGACGCGGTGAGCGCGCACCGGCGACCG 420  
 Db GCGCGCGGCTGCTCCATCGGACGTCGACGCGGTGAGCGCGCACCGGCGACCGGCGAC 5910  
 QY 421 CTGGCGGACCGGATGAGCGCGGCGGCTGCTGCGGCGGCTGCGGCGGCGGCGGCGGAG 480  
 Db CTGGCGGACCGGATGAGCGCGGCGGCTGCTGCGGCGGCTGCGGCGGCGGCGGCGGAG 5911  
 QY 481 CGGCTGTGGCTCGGGTCTGGTG 501  
 Db CGGCTGTGGCTCTGGGCTCTGGTG 5991

## RESULT 6

AAL61224/c

ID AAL61224 standard; DNA; 82746 BP.

XX

AC AAL61224;

XX

22-SEP-2003 (first entry)

XX

Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.

DE

XX

Maytansinoid; ansamitocin; antitumour; gene; ds.

XX

Actinosynnema pretiosum.

XX

WO2003045312-A2.

XX

05-JUN-2003.

XX

21-NOV-2002; 2002WO-US037547.

XX

21-NOV-2001; 2001US-0332158P.

XX

(UNIW ) UNIV WASHINGTON.

XX

Floss HG, Yu T, Leistner E;

XX

PI

XX

DR WPI; 2003-493374/46.  
 XX Novel maytansinoid produced by bacterial host cell transformed with  
 PT expression vector comprising open reading frame from ansamitocin gene  
 PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
 XX  
 PS Claim 7; Page 105-152; 160pp; English.  
 XX  
 CC The invention relates to maytansinoid produced by bacterial host cell  
 CC transformed with expression vector comprising open reading frame from  
 CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
 CC useful as a potent antitumour agent. The present sequence is A. pretiosum  
 CC ansamitocin ansamitocin biosynthetic gene cluster I  
 XX  
 SQ Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;

Query Match 64.9%; Score 325; DB 7; Length 82746;  
 Best Local Similarity 78.0%; Pred. No. 2.4e-42;  
 Matches 391; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
 QY 1 TCGTCTTCGCTGCTGCGTGCACATGGCGGTGCAGGCCCTGCAGCGCGCGAGTGTCTCC 60  
 Db TCGTCTTCGCTGCTGCGTGCACATGGCGGTGCAGGCCCTGCAGCGCGCGAGTGTCTCG 61710  
 QY 61 ATGGCGCTCGCGGCGGTGATGATGGGACCGTGCAGACGTTCTCGAGTTCCTCG 120  
 Db ATGGCGCTCGCGGCGGTGATGATGGGACCGTGCAGACGTTCTCGAGTTCCTCGAC 61591  
 QY 121 CGGCAGCGCGGCTGGCCCGGCGGTGCAGGCGTTCGCGACGCGCGGCGGACGCG 180  
 Db CGGCAGCGCGGCTGGCCCGGCGGTGCAGGCGTTCGCGACGCGCGGCGGACGCGC 61590  
 QY 181 ACCGGCTTGGTCCGAGGCGGTGCGGCTGCTCTGTTGAGGCGGTGCTCCGAGGCTCAGCGT 240  
 Db ACCGGCTTGGTCCGAGGCGGTGCGGCTGCTCTGTTGAGGCGGTGCTCCGAGGCTCAGCGT 61530  
 QY 241 CGCGGCGACCGGATGAGCGCGGCGGCTGCTGCGGCGGCTGCGGCGGCGGCGGCGGAG 300  
 Db CGCGGCGACCGGATGAGCGCGGCGGCTGCTGCGGCGGCTGCGGCGGCGGCGGAG 61470  
 QY 301 AAGCGGCTTGACGCGCGCGGCGGCGGCGTCCAGCGCGGTGATCCGCAAGGCACTGGCC 360  
 Db AAGCGGCTTGACGCGCGCGGCGGCGGCGTCCAGCGCGGTGATCCGCAAGGCACTGGCC 61410  
 QY 361 GCGCGCGGCTGCTCCATCGGACGTCGACGCGGTGAGCGCGCACCGGCGACCG 420  
 Db GCGCGCGGCTGCTCCATCGGACGTCGACGCGGTGAGCGCGCACCGGCGACCGGCGAC 61350  
 QY 421 CTGGCGGACCGGATGAGCGCGGCGGCTGCTGCGGCGGCTGCGGCGGCGGCGGCGGAG 480  
 Db CTGGCGGACCGGATGAGCGCGGCGGCTGCTGCGGCGGCTGCGGCGGCGGCGGAG 61290  
 QY 481 CGGCTGTGGCTCGGGTCTGGTG 501  
 Db CGGCTGTGGCTCTGGGCTCTGGTG 61230

## RESULT 7

AAD54645

ID AAD54645 standard; DNA; 113193 BP.

XX

AC AAD54645;

XX

26-JUN-2003 (first entry)

XX

Streptomyces nodosus amphotericin (amph) biosynthetic gene cluster.

XX

DE

XX

KW

XX

OS

XX

FH

XX

FT

CDS

Key

Location/Qualifiers

complement(4..1824)

```
FT FT /tag= a
FT FT /product= "ABC transporter encoded by S. nodosus amphG
FT FT gene"
FT CDS complement(1805..3628)
FT FT /tag= b
FT FT /product= "ABC transporter encoded by S. nodosus amphH
FT FT gene"
FT CDS 3840..4874
FT FT /tag= c
FT FT /product= "GDP-mannose dehydratase encoded by S. nodosus
FT FT amphiII gene"
FT CDS 5042..33574
FT FT /tag= d
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
FT FT nodosus amphiK gene"
FT CDS 33584..50518
FT FT /tag= e
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 15, 16 and 17 encoded by S. nodosus
FT FT amphiJ gene"
FT CDS 50571..56675
FT FT /tag= f
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 18 and thioesterase encoded by S.
FT FT nodosus amphiK gene"
FT CDS 56829..58019
FT FT /tag= g
FT FT /product= "Cytochrome P450 encoded by S. nodosus amphiL
FT FT gene"
FT CDS 58139..58648
FT FT /tag= h
FT FT /product= "ORF1, hypothetical protein"
FT CDS complement(58756..59610)
FT FT /tag= i
FT FT /product= "ORF2, hypothetical protein"
FT CDS 59869..61470
FT FT /tag= j
FT FT /product= "ORF3, hypothetical protein"
FT FT /transl_except= (pos:59869..59871, aa:Met)
FT CDS complement(61798..61995)
FT FT /tag= k
FT FT /product= "Ferredoxin encoded by S. nodosus amphiM gene"
FT CDS complement(62051..63250)
FT FT /tag= l
FT FT /product= "Cytochrome P450 encoded by S. nodosus amphiN
FT FT gene"
FT CDS complement(63250..64308)
FT FT /tag= m
FT FT /product= "NDP-sugar aminotransferase encoded by S.
FT FT nodosus amphiII gene"
FT CDS complement(64324..65775)
FT FT /tag= n
FT FT /product= "Glycosyl transferase encoded by S. nodosus
FT FT amphiDI gene"
FT FT /transl_except= (pos:65773..65775, aa:Met)
FT CDS 66081..70319
FT FT /tag= o
FT FT /product= "Polyketide synthase multienzyme housing
FT FT leading module encoded by S. nodosus amphiA gene"
FT CDS 70366..79938
FT FT /tag= p
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 1 and 2 encoded by S. nodosus by amphiB
FT FT gene"
FT CDS 79956..112709
FT FT /tag= q
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
FT FT nodosus by amphiC gene"
```

WO200297082-A2.

```
PD 05-DEC-2002.
XX
XX 27-MAY-2002; 2002WO-IE000071.
XX
XX 31-MAY-2001; 2001IE-00000527.
XX
XX (UYDU-) UNIV COLLEGE DUBLIN.
XX
XX Caffrey JP;
XX
XX WPI; 2003-201271/19.
DR P-PSDB; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
DR AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
DR AAE36129, AAE36130, AAE36131, AAE36132.
XX
XX Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
XX for preparing amphotericin derivative or analog antibiotic agent with
XX altered properties, in biosynthesis of polyketide other than
XX amphotericin.
XX
XX Claim 1; Page 52-114; 276pp; English.
XX
XX The invention relates to the gene cluster encoding the polypeptides
XX responsible for the biosynthesis of the polyene antibiotic amphotericin
XX (amph) of Streptomyces nodosus. Polynucleotides of the invention are
XX useful for preparing amphotericin derivatives or analogue antibiotic
XX agents with altered properties and in the biosynthesis of polyketides
XX other than amphotericin. amphiDII, amphiDII or amphiI mutants are useful
XX for producing amphotericin derivatives glycosylated with alternative
XX sugars; amphiDII or amphiDI gene sequences are useful in engineered
XX biosynthesis of peramanyl-amphoteronolide B; amphiDII or amphiDII and
XX amphiN gene sequences are useful in the engineered biosynthesis of
XX peramanyl-16-desacetyl-16-methyl amphoteronolide B; amphiDII, amphiDII
XX and amphiDI gene sequences are useful for preparing polypeptides capable
XX of addition of mycosamine to a polyketide other than amphoteronolide A or
XX B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
XX The present sequence is S. nodosus amphi biosynthetic gene cluster
XX
XX Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;
XX
XX Query Match 64.0%; Score 320.8; DB 7; Length 113193;
XX Best Local Similarity 77.6%; Pred. No. 1e-41;
XX Matches 388; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
QY 1 TGTCTTCCTGCTGCGCTGCACATGGCGGTGAGGCGCTGCAGCGCGCGAGTGTCTCC 60
Db 80568 TCTCTCTCTGCTCTCTCTCTGACCTGGCGCGCGAGGCGCTGCGCGCGAGTGTCTCC 80627
QY 61 ATGGCGCTCGCGGCGCGCTGATGTGATGGGACACGCTTCGAGACGTTCTGTGAGTTCTCG 120
Db 80628 CTGGCCCTTGGCGGCGCGCTGCTGATGGCGACGCCCGCTTCATCGCGTTCAAC 80687
QY 121 CGGACGCGGCGTGGCCCCCGACCGCGCTGCAGGCGCTTCGCGAGCGCGGACGCG 180
Db 80688 TGCAGAGCGGACTGGCGCGCGACGCGCGCTGCAAGCGCTTCTCCAGACGCGCGCG 80747
QY 181 ACCGCTGTGTCGAGGCGCTGCGGCTGCTGCTGAGCGGCTGTCGAGGCTCAGCGT 240
Db 80748 ACCGCTGTGCGCGAGGCGCGCGCTGTCCTGATGCGGCTGTCGAGCGCGCGCG 80807
QY 241 CGCGGCGCACAGGCTCTCTCGCGGTGCTCGCGGTTCAGCTCCGAGCGCGGTGCG 300
Db 80808 GAGGCGCGCAAGGCTCTCTCGCGGTACTGCGGCTCGCGGCTCAACGAGGCGCGCTCC 80867
QY 301 AACGCTTACGCGCGCGGACCGCGCGCTCCAGCAGCGGTGATTCGCGAGGCTGCGG 360
Db 80868 AACGCGCTGACCGCGCGCGGACCGCGCGCTCCGACAGCGGCTCATCGCGAGGCG 80927
QY 361 GCGCGCGGACTGTCACATCGGACGTCGAGCGCGGTGAGCGCGACGCGCGCGAGCG 420
Db 80928 CACGCGGCTGTCAGGCGGAGCGGATGTCGACGCGCTGAGGCGGATGCGGCGCGCG 80987
QY 421 CTGGCGGACCGCGATCGAGGCGCGGCGCTGCTGGCCACCTTACGCGCAGAACCGGAAACG 480
```

Db 80988 CTCGGTACCCGATCAGGCAAGGCGCTGTCGCCACCTACGGCCAGAACCGCGAGCGG 81047  
 QY 481 CCGCTGTGGCTCGGGTCGGT 500  
 Db 81048 CCGCTGTGCTCGGCTCCGT 81067  
 RESULT 8  
 AAL61170  
 ID AAL61171 standard; DNA; 9222 BP.  
 AC AAL61171;  
 XX 22-SEP-2003 (first entry)  
 DT Actinosynnema pretiosum polyketide synthase (PKS) gene #2.  
 DE Actinosynnema pretiosum; antitumour; polyketide synthase; enzyme; PKS;  
 XX Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;  
 KW gene; ds.  
 XX Actinosynnema pretiosum.  
 OS Actinosynnema pretiosum.  
 PN WO2003045312-A2.  
 XX 05-JUN-2003.  
 PD 21-NOV-2002; 2002WO-US037547.  
 XX 21-NOV-2001; 2001US-0332158P.  
 PR (UNIW ) UNIV WASHINGTON.  
 PA Floss HG, Yu T, Leistner E;  
 PI WPI; 2003-493374/46.  
 DR Novel maytansinoid produced by bacterial host cell transformed with  
 PT expression vector comprising open reading frame from ansamitocin gene  
 PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
 XX Disclosure; Page 53-58; 160pp; English.  
 XX The invention relates to maytansinoid produced by bacterial host cell  
 CC transformed with expression vector comprising open reading frame from  
 CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
 CC useful as a potent antitumour agent. The present sequence is A. pretiosum  
 CC ansamitocin gene cluster I polyketide synthase (PKS) gene  
 XX Sequence 9222 BP; 784 A; 3437 C; 3858 G; 1143 T; 0 U; 0 Other;  
 SQ Query Match 63.9%; Score 320.2; DB 7; Length 9222;  
 Best Local Similarity 78.5%; Pred. No. 1.9e-41;  
 Matches 398; Conservative 0; Mismatches 103; Indels 6; Gaps 1;  
 QY 1 TCGTCTTCGCTGTCGCTGCAATGCGGTGCGGCTCCAGCGGCGGCGGAGTGCTCC 60  
 Db 604 TCGTCTTCGCTGTCGCTGCAATGCGGTGCGGCTCCAGCGGCGGCGGAGTGCTCG 663  
 QY 61 ATGGCGCTCGCGGCGGCGGTGATGTCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
 Db 664 ATGGCGCTCGCGGCGGCGGCGGTGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 723  
 QY 121 CCGCAGCGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180  
 Db 724 CCGCAGCGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 783  
 QY 181 ACCGCGTGTGTCGAGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
 Db 784 ACCGCGTGTGTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 843  
 QY 241 CCGG 300

Db 844 GACGGCGCGCGGTGTTGGCGGTGTCGCGGCTGCGGCGGTGAACTCCGACGGCGGTCC 903  
 QY 301 AACGGCTTACG 360  
 Db 904 AACGGCTTACG 963  
 QY 361 GCGCGCGGACTGTGTCACATCGGACGTCGACGCGGTGAGCGCGCACCGGCGGCGGCGGCGG 420  
 Db 964 GACGGCGGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1023  
 QY 421 CTGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 474  
 Db 1024 CTGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1083  
 QY 475 GAAACGCGCGCTGTGCGGCTCGGGTCGGTG 501  
 Db 1084 GCGGGCGCGCTGTGCGGCTCGGGTCGGTG 1110  
 RESULT 9  
 AAL61170  
 ID AAL61170 standard; DNA; 14055 BP.  
 AC AAL61170;  
 XX 22-SEP-2003 (first entry)  
 DT Actinosynnema pretiosum polyketide synthase (PKS) gene #1.  
 DE Actinosynnema pretiosum; antitumour; polyketide synthase; enzyme; PKS;  
 KW gene; ds.  
 XX Actinosynnema pretiosum.  
 OS Actinosynnema pretiosum.  
 PN WO2003045312-A2.  
 XX 05-JUN-2003.  
 PD 21-NOV-2002; 2002WO-US037547.  
 XX 21-NOV-2001; 2001US-0332158P.  
 PR (UNIW ) UNIV WASHINGTON.  
 PA Floss HG, Yu T, Leistner E;  
 PI WPI; 2003-493374/46.  
 DR Novel maytansinoid produced by bacterial host cell transformed with  
 PT expression vector comprising open reading frame from ansamitocin gene  
 PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
 XX Disclosure; Page 45-53; 160pp; English.  
 XX The invention relates to maytansinoid produced by bacterial host cell  
 CC transformed with expression vector comprising open reading frame from  
 CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
 CC useful as a potent antitumour agent. The present sequence is A. pretiosum  
 CC ansamitocin gene cluster I polyketide synthase (PKS) gene  
 XX Sequence 14055 BP; 1304 A; 5585 C; 5518 G; 1648 T; 0 U; 0 Other;  
 SQ Query Match 63.9%; Score 320; DB 7; Length 14055;  
 Best Local Similarity 78.6%; Pred. No. 1.9e-41;  
 Matches 396; Conservative 0; Mismatches 105; Indels 3; Gaps 1;  
 QY 1 TCGTCTTCGCTGTCGCTGCAATGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60  
 Db 3550 TCGTCTTCGCTGTCGCTGCAATGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3609  
 QY 61 ATGGCGCTCGCGGCGGCGGCGGTGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 120  
 Db 3610 ATGGCGCTCGCGGCGGCGGCGGTGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 3669

QY 121 CGGACGCGCGGCTGGCCCGACGCGCGCTGCGAGGCGTTTCGCGACGCGCGGACGCGC 180  
 DB |||||  
 QY 3670 CGGCGGCGCGGCTGGCCCGACGCGCGCTGCGAGCCGTTTCGCGCTCGGCGGACGCGG 3729  
 DB |||||  
 QY 181 ACCGCGCTGGTCCGAGGCGGTGGGCTGCTCTCTGCTGAGCGCGCTGCTCCGAGGCTCAGCGT 240  
 DB |||||  
 QY 3730 ACGGCGTGGTCCGAGGCGCGCGGGTGGTGGTCTGCTGAGGCGGCTGCTCGAGCGCGCGCGC 3789  
 DB |||||  
 QY 241 CGCGGCGACAGGCTCTCGCGTGGTTCGCGGCTGCGCGGCTCACTCCGACGCGCGCTCG 300  
 DB |||||  
 QY 3790 AACGGGCGTGGGCTGCTCGCGGTGATCCGCGGCGACGCGGCTGAATCCGACGCGGCGCTCG 3849  
 DB |||||  
 QY 301 AACGCTTGAACGCGCGCCCGAACGCGCGCTGCCAGCAGCGCGTATCCGAAAGCACTGGCC 360  
 DB |||||  
 QY 3850 AACGGTTGACGCGCGCGGAGCGGTGGGCGGAGCGAGCGGCTGATCCGCGCGCGCTCGGG 3909  
 DB |||||  
 QY 361 GCCGCGGACTTCCACATCGGACGTCGACGCGGTGGAGGCGCACCGGACCGGGACGACC 420  
 DB |||||  
 QY 3910 GACGCGGCTCTGCTCACTTCGACGCTGGACGCGGTGGAGGCGCACCGGACCGGGACGCGG 3969  
 DB |||||  
 QY 421 CTGGGCGACCCATCGAGCCGAGCGGCTGCTGGCCACCTACGCGCACGACCGGCA--A 477  
 DB |||||  
 QY 3970 CTCGCTGATCCGATCGAGCGCGCGCGCTGGTTCGCGACCTACGCGCGGACCGGACCGGC 4029  
 DB |||||  
 QY 478 ACGCGCTGTGGCTCGGGTCTGGT 501  
 DB |||||  
 QY 4030 GAGCGCTGTGGCTCGGCTCGCTG 4053  
 DB |||||

## RESULT 10

AADI17184

ID AADI17184 standard; DNA; 65140 BP.

XX AADI17184;

AC 29-NOV-2001 (first entry)

XX Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.

XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
 KW antifungal; antibiotic; nys1; ds.

XX Streptomyces noursei.

PH Key Location/Qualifiers

CDS complement(l..1035)

FT /\*tag= a

FT /product= "NysD2 partial protein"

FT /note= "CDS does not include stop codon"

CDS complement(1056..2576)

FT /\*tag= b

FT /product= "NysD1 protein"

FT 2806..6906

FT /\*tag= c

FT /product= "NysA protein"

FT 6952..16530

FT /\*tag= d

FT /product= "NysB protein"

FT 16550..49840

FT /\*tag= e

FT /product= "NysC protein"

FT 50260..51015

FT /\*tag= f

FT /product= "NysE protein"

FT 51405..54305

FT /\*tag= g

FT /product= "NysR1 protein"

FT 54329..57190

FT /\*tag= h

FT /product= "NysR2 protein"

FT /note= "CDS does not include start codon"

FT 57180..59963

FT CDS

FT /\*tag= i  
 FT /product= "NysR3 protein"  
 FT 60415..61047  
 FT /\*tag= j  
 FT /product= "NysR4 (short) protein"  
 FT /note= "CDS does not include start codon"  
 FT 61736..62497  
 FT /\*tag= k  
 FT /product= "NysR5 protein"  
 FT /note= "CDS does not include start codon"  
 FT complement(62551..63615)  
 FT /\*tag= l  
 FT /product= "ORF2 protein"  
 FT /note= "CDS does not include start codon"  
 FT 63765..64961  
 FT /\*tag= m  
 FT /product= "ORF1 protein"

WO200159126-A2.

16-AUG-2001.

08-FEB-2001; 2001WO-GB000509.

08-FEB-2000; 2000GB-00002840.

10-APR-2000; 2000GB-00008786.

14-APR-2000; 2000GB-00009387.

(UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

(SNTF) SINTEF STIFTELSEN IND TEK FORSK.

(ALPH-) ALPHARMA AS.

(SINV-) SINVENT AS.

(DZIE-) DZIEGLEWSKA H.

(ZOTC-) ZOTCHEV S B.

(SEKU-) SEKUROVA O N.

(FUAE-) FJAEV E.

(BRAU-) BRAUTASET T.

(STRO-) STROM A R.

(VALL-) VALLA S.

Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX WPI; 2001-557614/62.

DR P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,

DR AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.

XX New nystatin polyketide synthase polynucleotides and polypeptides, useful

as antibiotics and antifungals.

XX Claim 2; Page 116-151; 266pp; English.

XX The present invention relates to the cloning and sequencing of the gene

cluster encoding a modular type I polyketide synthase (PKS) enzyme

involved in the biosynthesis of the macrolide antibiotic nystatin. The

nystatin PKS is useful as antifungal antibiotics. The present sequence is

a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster

SQ Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 U; 0 Other;

Query Match 63.4%; Score 317.6; DB 4; Length 65140;

Best Local Similarity 79.2%; Pred. No. 3.5e-41;

Matches 377; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTGGTGGCTGCACATGCGGTGCGAGGCGCTCGAGCGCGCGGCTCTCC 60

DB |||||

DB 44702 TCCTCTCCCTGGTGGCTGCACCTCCGCGAGCGCGCTTCGGCGCGGTGAGTGACG 44761

QY 61 ATGGGCGCTCGCGCGCGCGTGTGATGGACAGCGTTCGAGACGTTTCGTGAGTTCTCG 120

DB |||||

DB 44762 CTCGCGCTCGCGCGGTGGTGTGACGGTGTGATGTCACGCCAGGACGTTTCGTGAGTTCTCG 44821

QY 121 CGGCAGCGCGGGCTGGCCCGCGCGCTGCTGAAGGGGCTTCGCCGACGCGCGGACGCGC 180



```
FT misc_feature complement(31018. .32185)
FT /*tag= ag
FT /note= "AT3"
FT misc_feature complement(32185. .33439)
FT /*tag= ah
FT misc_feature complement(33505. .33715)
FT /*tag= ai
FT /*note= "ACP2"
FT misc_feature complement(33823. .34480)
FT /*tag= aj
FT /*note= "KR2"
FT misc_feature complement(34606. .35749)
FT /*tag= ak
FT misc_feature complement(35749. .37144)
FT /*tag= am
FT /*note= "AT2"
FT misc_feature complement(37145. .38296)
FT /*tag= an
FT /*note= "KS2"
FT misc_feature complement(38371. .38581)
FT /*tag= ao
FT /*note= "ACP1"
FT misc_feature complement(38677. .39307)
FT /*tag= ap
FT /*note= "KR1"
FT misc_feature complement(39442. .40609)
FT /*tag= aq
FT /*note= "DH1"
FT misc_feature complement(40609. .41842)
FT /*tag= ar
FT /*note= "AT1"
FT misc_feature complement(41842. .43093)
FT /*tag= as
FT /*note= "KS of extender module 1 (KS1)"
FT misc_feature complement(43144. .43660)
FT /*tag= at
FT /*note= "ACP of loading domain"
FT misc_feature complement(43777. .44629)
FT /*tag= au
FT /*note= "ER of loading domain"
FT misc_feature complement(44974. .46573)
FT /*tag= av
FT /*note= "CoA ligase of loading domain"
FT CDS 46754. 47788
FT /*tag= aw
FT /*note= "fkb gene"
FT CDS 47785. .52272
FT /*tag= ax
FT /*note= "fkb gene"
FT CDS 52275. 71465
FT /*tag= ay
FT /*note= "fkb gene"
FT misc_feature 52362. .53576
FT /*tag= az
FT /*note= "KS7"
FT misc_feature 53577. 54716
FT /*tag= bb
FT /*note= "AT7"
FT misc_feature 54717. .55871
FT /*tag= ba
FT /*note= "PH7"
FT misc_feature 56019. .56819
FT /*tag= bc
FT /*note= "ER7"
FT misc_feature 56943. .57575
FT /*tag= bd
FT /*note= "KR7"
FT misc_feature 57710. .57920
FT /*tag= be
FT /*note= "ACP7"
FT misc_feature 57990. .59243
FT /*tag= bf
```

```
FT misc_feature /note= "KS8"
FT 59244. .60398
FT /*tag= bg
FT misc_feature /note= "AT8"
FT 60399. .61412
FT /*tag= bh
FT /*note= "DH8 (inactive)"

Query Match 62.8%; Score 314.4; DB 3; Length 77536;
Best Local Similarity 76.8%; Pred. No. le-40; Mismatches 116; Indels 0; Gaps 0;
Matches 384; Conservative 0;

QY 1 TCGTCTTCGTCGCGCTGCACATGGCGGTGCAGGCCCTGCAGCGCGGAGTGTCC 60
DB 42592 TCGTCGTCGTCGCGCTGCACAGGCGGCACAGGCACTCGCGCGGAGTGTCC 42533
QY 61 ATGGCGCTCGCGCGGCGGTGATGATGATGGGACGCTGAGAGAGCTTGTGAGTTCG 120
DB 42532 CTGGCGCTGTCGCGCGGTGTCACGTCATGCGCAACACCGAGAGCTCTGAGTTC 42473
QY 121 CGGCAGCGCGGCTGGCCCGCGCGCTGCAAGGCTTCGCGCGAGCGCGGACGCG 180
DB 42472 CGGCAGCGCGGCTTCCTCCCGACGACGATGAGGCTTCGCGCGAGCGCGG 42413
QY 181 ACCGCTGTCGCGGCGGTGCGGCTGCTCTGCTGAGCGGCTGTGCGAGGCTCAG 240
DB 42412 ACCGCTGTCGCGGCGGTGCGGCTTCTCTGCTGAGCGGCTCTCGACGCGG 42353
QY 241 CGCGGCGACAGGTCCTCGCGGTGTCGCGGCTGCGGCTCACTCCAGCGCGGTCG 300
DB 42352 AACGCGCACACCGCTTCTCGCGGTGTCGCGGCTGCGGCGTGAACCGAGAG 42293
QY 301 AACGCTTGCAGCGCGCGCGCGCTCCAGAGCGCGCTGATCCGCAAGGCACTGG 360
DB 42292 AACGACTGTGCGCGCGCGCGCGCTGCGCGAGCGGCTGCGCGAGCGGCTG 42233
QY 361 GCCTCGGCTGTCCACATCGGACGTCGACGCGGTGAGGCGCGACCGCACCGGAG 420
DB 42232 AACGCGGACTGCGCGCGCGCGCTGTCGAGCGTGGTGGAGCGCGACCGG 42173
QY 421 CTGGCGGACCGGATCGAGCGCGCGCTGTCGCGCACTAGCGCGAGACCGGAA 480
DB 42172 CTGGCGGACCGGATCGAGCGCGCGCTGTCGCGCGCTAGCGGCGAGCGG 42113
QY 481 CGCTGTGCTGCGGTGCGGT 500
DB 42112 CGCTCTTACTCGGTTCGCT 42093

RESULT 12
AAZ87283
ID AAZ87283 standard; DNA; 15872 BP.
XX
AC AAZ87283;
XX
DT 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
KW Desamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae; ATCC15439.
XX
FH Key Location/Qualifiers
FT CDS 20..13912
FT /*tag= a
FT /product= "vep ORF 1 amino acid sequence #1 (AAZ77177)"
FT CDS 14056..14151
```



```
FT      /*tag= b
FT      /product= "vcp ORF 1 amino acid sequence #3 (AAV77199)"
FT      14167..15827
FT      /*tag= c
FT      /product= "vcp ORF 1 amino acid sequence #2 (AAV77178)"
XX
XX WO200000620-A2.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99WO-US014398.
XX
XX 26-JUN-1998; 98US-00105537.
XX
XX (MINU ) UNIV MINNESOTA.
XX
XX Sherman DH, Liu H, Xue Y, Zhao L;
XX
XX WPI; 2000-160679/14.
XX
XX P-PSDB; AAV77177, AAV77178, AAV77199.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
XX synthesis of methymycin and pikromycin.
XX
XX Example 3; Fig 23; 438pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid segment
XX comprising a desosamine biosynthetic gene cluster, a fragment or its
XX biologically active variant, where the nucleic acid sequence is not
XX derived from the eryc gene cluster of Saccharopolyspora erythraea or
XX Streptomyces antibioticus. The invention also relates to a macrolide
XX biosynthetic gene cluster, or fragments thereof. The macrolide
XX biosynthetic gene cluster encodes proteins which synthesize methymycin,
XX pikromycin, neomethymycin, narbomycin or a combination of these
XX compounds. Recombinant or augmented cells comprising the desosamine
XX and/or macrolide biosynthetic gene clusters are useful for the production
XX of biologically active macrolides. The macrolide biosynthetic proteins
XX are useful for synthesis of methymycin, pikromycin, neomethymycin and
XX narbomycin. The alternative termination of polyketide synthesis may be
XX useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
XX monomers. The compounds produced by the recombinant host cells are useful
XX as biopolymers, e.g., in packaging or biomedical applications, to
XX engineer PHA monomer synthases or to prepare biologically active agents,
XX such as chemotherapeutics, immunosuppressants, agents to treat asthma,
XX chronic obstructive pulmonary disease as well as other diseases involving
XX respiratory inflammation, cholesterol-lowering agents or macrolide-based
XX antibiotics which are active against a variety of organisms, e.g.,
XX bacteria, including multi-drug resistant pneumococci and other
XX respiratory pathogens, as well as viral parasitic pathogens, or as crop
XX protection agents (e.g., fungicides or insecticides) via expression of
XX polyketides in plants. The present sequence represents a Streptomyces
XX venezuelae ATCC 15439 DNA sequence, designated vcp ORF 1 in the
XX specification, which actually contains 3 open reading frames, which
XX encode proteins AAV77177-Y77178 and AAV77199. The vcp ORF 1 protein is
XX defined in the specification as a PHA monomer synthase. (Updated on 15-
XX SEP-2003 to standardise OS field)
XX
XX
XX Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 U; 0 Other;
XX
XX Query Match 61.4%; Score 307.4; DB 3; Length 15872;
XX Best Local Similarity 76.9%; Pred. NO. 1.7e-39;
XX Matches 390; Conservative 0; Mismatches 111; Indels 6; Gaps 1;
XX
XX 1 TCCTCTTCGCTGCTCGGCTGCATCGAGTGTGATGCGACGGTTCGAGTTCCTCC 60
XX |||||
XX 9017 TCCTCTTCGCTGCTCGGCTGCATCGAGTGTGATGCGACGGTTCGAGTTCCTCC 9076
XX |||||
XX 61 ATGGCGCTCGCGGGGGGGTGTGATGCGTGTGATGCGACGGTTCGAGTTCCTCG 120
XX |||||
XX 9077 CTCGCGCTGCGGGGGGGGTGCGGTGTGATGCCGATCCGGCGGCTTCGTGAGTTCCTCC 9136
XX |||||
XX 121 CGGCAGCGCGGCTGCGCCCGGACCGCGCTGCAAGGGGTTTCGCGACGGCGCGACGGC 180
XX |||||
```

```
Db      9137 CGGCAGAGGGGCTGCGCCCGCAGCGCCGCTGCAAGCGCTTCGCGCGCGCGCGGC 9196
QY      181 ACCGGCTGCTCCGAGGGCGTCCGGCTGCTCTTGGTGGAGCGGCTGTCCGAGGTCAGCGT 240
|||
Db      9197 ACCGGTGGCGGAGGGCGTCCGGCTGCTCTTGGAGCGGCTGTCCGAGCGCGCGC 9256
|||
QY      241 CGCGGCACACAGTCTCTCGCGCTGCTCCGGCTGCTCCGGCTGCTCAACTCCGAGCGGCTCG 300
|||
Db      9257 GCGGGGCACACAGTCTCTCGCGCTGCTCCGGCTGCTCAACTCCGAGCGGCTCGCTCC 9316
|||
QY      301 AACGGCTTACAGCGCCCGGACGCGCTCCAGCAGCGGCTGATCCGCAAGGCACTTGCC 360
|||
Db      9317 AACGGCTTACAGCGCCCGGACGCGCTCCAGCAGCGGCTGATCCGCAAGGCACTTGCC 9376
|||
QY      361 GCGCGCGACTGTCCACATCGGACGCTCGAGCGGCTGGAGCGCGCACCGGACGAGACC 420
|||
Db      9377 GACGCGGGCTGTCCCGGAGGAGCTGGAGCGGCTCGAGCGCGCACCGGACGAGACC 9436
|||
QY      421 CTGGCGACCCCGATCGAGCGGAGCGGCTGCTCGGCACTTACCGCGCAGAACCC-----GG 474
|||
Db      9437 CTGGCGACCCCGATCGAGCGGAGCGGCTGCTCGGCACTTACCGCGCAGAACCC-----GG 9496
|||
QY      475 GAAACCGCGCTGCGTCCGGTCCGGTCCGGT 501
|||
Db      9497 GACCACCGCTGTGGCTCGGCTCGCTG 9523
|||

RESULT 13
AAT80413
ID      AAT80413 standard; DNA; 43280 BP.
XX
XX AAT80413;
XX
DT      27-FEB-1998 (first entry)
XX
DE      Tylactone synthase gene cluster.
XX
KW      Tylactone synthase gene cluster; tylG gene; multifunctional protein;
XX polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS      Streptomyces fradiae.
XX
XX Key Location/Qualifiers
XX CDS 816..14243
|||
XX /*tag= a
XX /transl_except= (pos: 816..818, aa: Met)
XX /note= "ORF1 encodes protein shown in AAW22601"
XX 14351..19945
|||
XX /*tag= b
XX /transl_except= (pos: 14351..14353, aa: Met)
XX /note= "ORF2 encodes protein shown in AAW22602"
XX 20010..31199
|||
XX /*tag= c
XX /transl_except= (pos: 20010..20012, aa: Met)
XX /note= "ORF3 encodes protein shown in AAW22603"
XX 31232..36067
|||
XX /*tag= d
XX /note= "ORF4 encodes protein shown in AAW22604"
XX 36249..41774
|||
XX /*tag= e
XX /note= "ORF5 encodes protein shown in AAW22605"
XX

PN      EP791655-A2.
XX
PD      27-AUG-1997.
XX
PF      19-FEB-1997; 97EP-00301056.
XX
PR      22-FEB-1996; 96US-0012078P.
XX
PA      (ELIL ) LILLY & CO ELI.
XX
PI      Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;
```









Qy	421	CTGGGGACCCGATCAGAGCGCGAGGGCGTCTCTGGCGACCTACGGCCGGAACC-----GG	474
Db	9437	CTCGGGACCCCATCAGAGCGCGGGGGCGTCTCTCGCGCCTCCGACGGAACCTTCCGGC	9496
Qy	475	GAACAACCCGCTGTGGCTCGGGTCGGTG	501
Db	9497	GACACACCGCTGTGGCTCGGCTCGCTG	9523

```

RESULT 4
US-09-091-609-3
; Sequence 3, Application US/09091609
; Patent No. 6600029
; GENERAL INFORMATION:
; APPLICANT: SHERMAN, DAVID H.
; APPLICANT: WILLIAMS, MARK D.
; APPLICANT: XUE, YONGQUAN
; TITLE OF INVENTION: METABOLIC ENGINEERING OF
; TITLE OF INVENTION: POLYHYDROXYALKANOATE MONOMER SYNTHASES
; FILE REFERENCE: 600.297US2
; CURRENT APPLICATION NUMBER: US/09/091,609
; CURRENT FILING DATE: 1998-06-19
; EARLIER APPLICATION NUMBER: PCT/US96/20119
; EARLIER FILING DATE: 1996-12-18
; EARLIER APPLICATION NUMBER: 60/008,847
; EARLIER FILING DATE: 1995-12-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (14148) ... (15824)
US-09-091-609-3

```

Query Match	51.4%	Score 307.4	DB 4	Length 15872	
Best Local Similarity	76.9%	Pred. No. 2e-47			
Matches 390	Conservative	0	Mismatches 111	Indels 6	Gaps 1
Qy	1	TCGCTTCGCTGCTCGCGTGCACATCGCGCGTGCAGGCCCTGCAGCGCGCGAGTGCCTCC	60		
Db	9017	TCCTTCCTCGCTGTGTGGCGCTGCATCTGCGCGGTATCGGGCGTGTGCGGCACGCGAGTGCGGG	9076		
Qy	61	ATGGCGCTTCGCGGGCGCGCTGATGGTGATCGGGACCGGTTCGAGACGTTTCGTCGAGTTCTCG	120		
Db	9077	CTCGCGTTCGCGGGCGGGGTGGCGGTGATGGCCGATCCGGGGCGTTCTGTGAGTTCTCC	9136		
Qy	121	CGGCAGCGCGGGGCTGGCCCCCGACGGGCCGTGCAAGCGTTTCGCCGACGCGCGGACGGC	180		
Db	9137	CGGCAGAAAGGGCTTGGCGCGCGACGGCGCGTGTCAAAGCGTTCTCGGCCGCGCGACGGC	9196		
Qy	181	ACCGGCTGTTCCGAGGGGCTCGGGCTGCTCTGTGTGAGCGGCTGTCTCCGAGGCTCAGCGT	240		
Db	9197	ACCGGCTGGGCGGAGGGCGTTCGGCGTGTCTGCTTCGAGGGCGGTGTTCGACGCGCGCGCGC	9256		
Qy	241	CGCGGGACACAGGTCCTTCGCGCTGTTCGGGGTTCGGCGGTCAACTCCGACGCGCGGTGCG	300		
Db	9257	CGCGGGACACACGGTCTCTCGGCTTGTTCACCGGCAACCGCGGTCAACACGAGACGGTGCCTCC	9316		
Qy	301	AACGGCTTGACGGCCCCGAAACGGGCCGCTCCACGACGCGCGTGTATCCGCAAGGCACATGGCC	360		
Db	9317	AACGGGCTGACCGCGGCCAACGGGCCACGCCACAGCAAACGCGTCACTGCGGAGGCGCTGCC	9376		
Qy	361	GCCGCGGACTGTCTCACATCGACGAGCTCGACCGGTGGAGGCGCACGCGACCGGGACGACC	420		
Db	9377	GACGCGGGCTGTCCCCGGAGGACGTGGACCGGTTCGAGGCGCACGCGACCGGCACCCGG	9436		
Qy	421	CTGGGCGACCCGATCGAGGCCGAGGGCGTGTGGCCACCTTACGGCCAGAAC	474		
Db	9437	CTGGGCGACCCCATCGAGGCCGGGGCGTGTCTGCGCGCTTCGGACGCGAACCGTGTCCGGC	9496		

```

Qy 475 GAAACCGCGCTGCTGGCTCGGCTCGCTG 501
Db 9497 GACCACCGCTGCTGGCTCGGCTCGCTG 9523

```

```

RESULT 5
US-08-804-227C-1
; Sequence 1, Application US/0804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kustoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..35067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
US-08-804-227C-1

```

	Query Match	61.4%	Score 307.4;	DB 2;	Length 43280;
	Best Local Similarity	76.9%;	Pred. No. 1.9e-47;		
	Matches 390;	Conservative 0;	Mismatches 111;	Indels 6;	Gaps 1;
QY	1	TCGTCCTGCTGGTCGCGCTGCACATGCGCGGTGCAGCCCTTCGACGCGCGGAGTGTCTCC	60		
Db	36879	TCCTCTCCTCTGGTGGCGCTGCATCTGGCGGTACGGCGCTGCGGCACGCGGAGTGCGGG	36938		
QY	61	ATGCGCGCTCGCGGGCGCGCTGTGATGGTGTATGGGACAGCGTTCGATCGAGTTCCTCG	120		
Db	36939	CTCGCGCTTGGCGGCGCGGGTGGCGGCTGATGGCCGATCCGCGCGCGTTCGTGGAGTTCCTCC	36998		



QY 121 CGGAGCGGGGTGGCCCGCCGACGCGCGCTGCAAGGGCTTCGCCGACGCGCGGACGCGC 180  
Db |||||  
QY 36999 CGGCAAGGGGTGGCCCGCCGACGCGCGCTGCAAGGGCTTCGCCGCGCGGACGCGC 37058  
Db |||||  
QY 181 ACCGGTGTGTCGAGGGCGTCGGGCTGCTCTCGGTGAGCGGTGTCGAGGCTCAGCGT 240  
Db |||||  
QY 37059 ACCGGTGTGTCGAGGGCGTCGGGCTGCTCTCGGTGAGCGGTGTCGAGGCTCAGCGT 37118  
Db |||||  
QY 241 CGGCGGACACAGTCTCTCGCGTGTGTCGCGGTGCGCGGTCAACTCCGACGCGCGGTGCG 300  
Db |||||  
QY 37119 GCGGGGACACAGTCTCTCGCGTGTGTCGCGGTGCGCGGTCAACTCCGACGCGGTGCG 37178  
Db |||||  
QY 301 AACGGTGTGACGCGCGCGGCGCGCGTCCGAGCGGTGATCCGCAAGGCACTGGCC 360  
Db |||||  
QY 37179 AACGGTGTGACGCGCGCGGCGCGGTCCGAGCGGTGATCCGCAAGGCACTGGCC 37238  
Db |||||  
QY 361 GCGCGCGGACTGTCCACATCGACGTCGACGCGGTGAGCGCGACGCGCACCGGACGACC 420  
Db |||||  
QY 37239 GACGCGGGCTGTCCCGCGAGGACGTGGAGCGGTGAGCGCGACGCGCACCGGACCGCG 37298  
Db |||||  
QY 421 CTGGGCGACCCGATCGAGCGCGGCGGTGCTGGCCACCTACGCGCACGACCC-----GG 474  
Db |||||  
QY 37299 CTGGGCGACCCGATCGAGCGCGGCGGTGCTGGCCGCGCTCGCGCGCTCGGACGGAACGTTCCGCGC 37358  
Db |||||  
QY 475 GAAACGCGCGCTGTGGCTCGGCTCGGTG 501  
Db |||||  
QY 37359 GACCACCGCTGTGGCTCGGCTCGGTG 37385  
Db |||||

## RESULT 6

US-09-320-878-19  
; Sequence 19, Application US/09320878A  
; Patent No. 6117659

; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.

; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120

; CURRENT APPLICATION NUMBER: US/09/320,878A  
; CURRENT FILING DATE: 1999-05-27

; EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
; EARLIER FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: 60/119,139  
; EARLIER FILING DATE: 1999-02-08

; EARLIER APPLICATION NUMBER: 60/100,880  
; EARLIER FILING DATE: 1998-09-22

; EARLIER APPLICATION NUMBER: 60/087,080  
; EARLIER FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 19  
; LENGTH: 38506

; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae

US-09-320-878-19

Query Match 61.0%; Score 305.8; DB 3; Length 38506;  
Best Local Similarity 76.7%; Pred. No. 3.8e-47;

Matches 389; Conservative 0; Mismatches 112; Indels 6; Gaps 1;

QY 1 TCGTCTTCGCTGCTGCGGTGACATGCGGCTTCC 60  
Db |||||  
QY 3727 TCGTCTGCTGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTTCC 3786  
Db |||||  
QY 61 ATGGCGCTCGCGGCGCGGTGATGGTGGGACGCTCGAGACGCTTCGTCGAGTTCTCG 120  
Db |||||

Db 3787 ATGGCGCTCGCGGCGCGGTGATGGTGGGACGCTGATGCCACGCGCGGATGTTCTCGAGTTTCAGC 3846  
QY |||||  
QY 121 CGGCAAGGGGTGGCCCGCCGACGCGCGCTGCAAGGGCTTCGCCGACGCGCGGACGCGC 180  
Db |||||  
QY 3847 CGGCAAGGGGTGGCCCGCCGACGCGCGCTGCAAGGGCTTCGCCGACGCGCGGACGCGC 3906  
Db |||||  
QY 181 ACCGGTGTGTCGAGGGCGTCGGGCTGCTCTCGGTGAGCGGTGTCGAGGCTCAGCGT 240  
Db |||||  
QY 3907 ACCAGTGTGTCGAGGGCGTCGGGCTGCTCTCGGTGAGCGGTGTCGAGGCTCAGCGT 3966  
Db |||||  
QY 241 CGGCGGACACAGTCTCTCGCGTGTGTCGCGGTGCGCGGTCAACTCCGACGCGCGGTGCG 300  
Db |||||  
QY 3967 AACGGACACAGTCTCTCGGCTGCTCGCGCGGCGCGGTGAAACGAGACGCGCGGACG 4026  
QY 301 AACGGTGTGACGCGCGCGGCGCGGTCCGAGCGGTGATCCGCAAGGCACTGGCC 360  
Db |||||  
QY 4027 AACGGCTCAGCGCTCCGAAACGCGGCTCGGACGCGGTGATCCGCGCGCGGTGCGG 4086  
QY 361 GCGCGCGGACTGTCCACATCGGACGTCGACGCGGTGAGCGCGACGCGCACCGGACGACC 420  
Db |||||  
QY 4087 GACGCGGGCTGACGACCTCCGACGCTGAGCGCTGTCGAGCGCACGCGCACGCGCACGCGCA 4146  
QY 421 CTGGGCGACCCGATCGAGCGCGGCGGTGCTGGCCACCTACGCGCGACGACCC-----G 474  
Db |||||  
QY 4147 CTGGGCGACCCGATCGAGCGCGGCGGCTGATCGCCACCTACGCGCGCGGTGAGCGAC 4206  
QY 475 GAAACGCGCGCTGTGGCTCGGCTCGGTG 501  
Db |||||  
QY 4207 GAACAGCGCGCTGCGCTCGGCTCGGTG 4233  
Db |||||

## RESULT 7

US-09-141-908-1

; Sequence 1, Application US/09141908  
; Patent No. 6503741

; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary

; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li

; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a  
; FILE REFERENCE: 300622002100

; CURRENT APPLICATION NUMBER: US/09/141,908  
; CURRENT FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: PROV. 60/076,919  
; EARLIER FILING DATE: 1998-03-05

; EARLIER APPLICATION NUMBER: PROV. 60/087,080  
; EARLIER FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1  
; LENGTH: 38506

; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae

US-09-141-908-1

Query Match 61.0%; Score 305.8; DB 4; Length 38506;  
Best Local Similarity 76.7%; Pred. No. 3.8e-47;

Matches 389; Conservative 0; Mismatches 112; Indels 6; Gaps 1;

QY 1 TCGTCTTCGCTGCTGCGGTGACATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTTCC 60  
Db |||||  
QY 3727 TCGTCTGCTGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTTCC 3786  
Db |||||  
QY 61 ATGGCGCTCGCGGCGCGGTGATGGTGGGACGCTCGAGACGCTTCGTCGAGTTCTCG 120  
Db |||||



QY 121 CGGAGCGCGGGCTGGCCCCCGACGGCGCGCTCGAAGGCGTTCCCGCAGCGGCGCGGACGGC 180  
DB 1092 CGGAGCGCGGGCTGGCCCCCGACGGCGCGCTCGAAGGCGTTCCCGCAGCGGCGCGGACGGC 1151  
QY 181 ACCGCTGTGTCGAGGGGCTGGGGCTGCTCTGCTGGTGGAGCGGCTGTCCGAGGCTCAGCGT 240  
DB 1152 ACGAGCTTCCGAGGGGCGCGGCTGGCTTCTGCTGGTGGAGCGGCTGTCCGAGGCTCAGCGT 1211  
QY 241 CGCGGCGACCAAGGCTCTCTCGCGCTGGTTCGCGGCTCGGCGGTCAACTCCGACGCGCGGTG 300  
DB 1212 CACGGCGCACCGCTCTCTCGCGCTGGTTCGCGGCTCGGCGGTCAACTCCGACGCGCGGTG 1271  
QY 301 AACGCTTTCAGCGCGCGCGCGCGCGCTCCGAGCGCGGCTGATCCGCAAGGCACTGGGC 360  
DB 1272 AACGCTTTCAGCGCGCGCGCGCGCGCTCCGAGCGCGGCTGATCCGCAAGGCACTGGGC 1331  
QY 361 GCGCGCGGACTCTCCACATCGACGCTCGACGCGCGGTGGAGCGCGCACCGGACCGGACGAC 420  
DB 1332 AACGGGAACCTCACCGCGCGCTGTCGACGCGGTGGAGCGCGCACCGGACCGGACGAC 1391  
QY 421 CTGGGCGACCGGATCGAGCGCGGCGCTGCTGGCGCACCTACGCGCAGAACCGGGAACG 480  
DB 1392 CTCGGCGACCGGATCGAGCGCGGCGCTGCTGGCGCACCTACGCGCAGAACCGGGAACG 1451  
QY 481 CGCTGTGGCTCGGGTCGGTG 501  
DB 1452 CCCCTGCTGCTCGGTCGCTG 1472

## RESULT 10

US-09-410-551B-24  
; Sequence 24, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; CURRENT FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 24  
; LENGTH: 4725  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)..(4724)  
US-09-410-551B-24

Query Match 60.7%; Score 304.2; DB 4; Length 4725;  
Best Local Similarity 75.4%; Pred. No. 7.8e-47;  
Matches 378; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTGGCTGCACATGGCGGTGCGAGCCCTGCGAGCGCGGAGTGTCC 60  
DB 972 TCGTCTTCGCTGGCTGCACATGGCGGTGCGAGCCCTGCGAGCGCGGAGTGTCC 1031

QY 61 ATGGCGCTCGCGGGCGGGCTGATGCTGATGGCAGCGGTGCGACCGTTCGTTCAGTTCTCG 120  
DB 1032 CTCGCCCTGGTTCGGCGGGTTCACCGTTCATGGCGTCCGCCGCGGATTCGTTCAGTTCTCC 1091  
QY 121 CGGAGCGCGGGCTGGCCCCCGACGGCGCGCTCGAAGGCGTTCCCGCAGCGGCGCGGACGGC 180  
DB 1092 CGGAGCGCGGGCTGGCCCCCGACGGCGCGCTCGAAGGCGTTCCCGCAGCGGCGCGGACGGC 1151  
QY 181 ACCGCTGTGTCGAGGGGCTGGGGCTGCTCTGCTGGTGGAGCGGCTGTCCGAGGCTCAGCGT 240  
DB 1152 ACGAGCTTCCGAGGGGCGCGGCTGGCTTCTGCTGGTGGAGCGGCTGTCCGAGGCTCAGCGT 1211  
QY 241 CGCGGCGACCAAGGCTCTCTCGCGCTGGTTCGCGGCTCGGCGGTCAACTCCGACGCGCGGTG 300  
DB 1212 CACGGCGCACCGCTCTCTCGCGCTGGTTCGCGGCTCGGCGGTCAACTCCGACGCGCGGTG 1271  
QY 301 AACGCTTTCAGCGCGCGCGCGCGCTCCGAGCGCGGCTGATCCGCAAGGCACTGGGC 360  
DB 1272 AACGCTTTCAGCGCGCGCGCGCGCTCCGAGCGCGGCTGATCCGCAAGGCACTGGGC 1331  
QY 361 GCGCGCGGACTCTCCACATCGACGCTCGACGCGCGGTGGAGCGCGCACCGGACCGGACGAC 420  
DB 1332 AACGGGAACCTCACCGCGCGCTGTCGACGCGGTGGAGCGCGCACCGGACCGGACGAC 1391  
QY 421 CTGGGCGACCGGATCGAGCGCGGCGCTGCTGGCGCACCTACGCGCAGAACCGGGAACG 480  
DB 1392 CTCGGCGACCGGATCGAGCGCGGCGCTGCTGGCGCACCTACGCGCAGAACCGGGAACG 1451  
QY 481 CGCTGTGGCTCGGGTCGGTG 501  
DB 1452 CCCCTGCTGCTCGGTCGCTG 1472

## RESULT 11

US-09-410-551B-30  
; Sequence 30, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; CURRENT FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 4737  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)..(4736)  
US-09-410-551B-30

Query Match 60.7%; Score 304.2; DB 4; Length 4737;  
Best Local Similarity 75.4%; Pred. No. 7.8e-47;  
Matches 378; Conservative 0; Mismatches 123; Indels 0; Gaps 0;



```
; NAME/KEY: CDS
; LOCATION: (3)..(4817)
US-09-410-551B-32

Query Match      50.7%; Score 304.2; DB 4; Length 4818;
Best Local Similarity 75.4%; Pred. No. 7.8e-47;
Matches 378; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 1 TCCTCTTCGCTGCTGCGCTGCACATCGGCGGTGCAGCGCTCGAGCGCGGAGTCTCC 60
Db 972 TCCTCTTCGCTGCTGCGCTGCACATCGGCGGTGCAGCGCTCGAGCGCGGAGTCTCC 1031

QY 61 ATGCGCTCCGCGCGCGCTGATGGTATGGGACACGCTCGAGAGTTCGTCG 120
Db 1032 CTCGCCCTGCTGCGCGGTGTCAGGTCATGCGCTCGCGCGGATTCGTCGAGTCTCC 1091

QY 121 CGGACGCGCGGCTGGCGCGCGCGCGCTGCAAGCGGTTCGCGCGCGGACGCG 180
Db 1092 CGGACGCGCGGCTGGCGCGCGCGCGCTGCAAGCGGTTCGCGCGCGGACGCGT 1151

QY 181 ACCGGCTGGTCCGAGGCGCTGCGCTCTCTGTTGAGCGGCTCGAGGCTCAGCGT 240
Db 1152 ACGAGCTTCGCGAGGCGCGCTGCGCTCTCTGTTGAGCGGCTCGAGGCTCAGCGT 1211

QY 241 CGCGGCGACACAGTCTCGCGGTGCTCGCGGTTCGCGGTTCGCGGTTCGCGGT 300
Db 1212 CACGGCGACACAGTCTCGCGGTGCTCGCGGTTCGCGGTTCGCGGTTCGCGGT 1271

QY 301 AAGCGCTTGAGCGCGCGCGCGCGCGCGCGCTCCAGAGCGGTGATCCGCAAGGCACTGGCC 360
Db 1272 AACGGCTTGAGCGCGCGCGCGCGCGCGCGCTCCAGAGCGGTGATCCGCAAGGCACTGGCC 1331

QY 361 GCCCGCGAGTCTCCATCGGAGCTCGAGCGGTGAGCGCGCGCGCGCGCGCGCGCG 420
Db 1332 AAGCGAACTACCGCGCGGTGATGCGCGGTGAGCGCGCGCGCGCGCGCGCGCG 1391

QY 421 CTGGGCGACCGATCGAGCGCGCGCGCTGCTGCGCACCTACGCGCGCGCGCGCGCGCG 480
Db 1392 CTGGGCGACCGATCGAGCGCGCGCGCTGCTGCGCACCTACGCGCGCGCGCGCGCGCG 1451

QY 481 CCCTCTGCTCGGCTCGGCTCGG 501
Db 1452 CCCCTGCTCGGCTCGGCTCGG 1472

RESULT 14
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

Query Match      60.7%; Score 304.2; DB 3; Length 13842;
Best Local Similarity 76.5%; Pred. No. 7.6e-47;
Matches 388; Conservative 0; Mismatches 113; Indels 6; Gaps 1;

QY 1 TCCTCTTCGCTGCTGCGCTGCACATGGCGGTGCGCGCTCGAGCGCGGAGTCTCC 60
Db 3844 TCCTCTTCGCTGCTGCGCTGCACATGGCGGTGCGCGCTCGAGCGCGGAGTCTCC 3903
```

```
QY 61 ATGCGCTTCGCGCGCGCTGATGGTATGGGACACGCTCGAGAGTTCGTCGAGTTCCTCG 120
Db 3904 ATGCGCTTCGCGCGCGCTGATGGTATGGGACACGCTCGAGAGTTCGTCGAGTTCAGC 3963

QY 121 CGGACGCGCGGCTGGCGCGCGCGCTCAAGCGGTTCGCGCGCGGACGCGCGCGCG 180
Db 3964 CGGACGCGCGGCTGGCGCGCGCGCTCAAGCGGTTCGCGCGCGGACGCGCGCGCG 4023

QY 181 ACCGGCTGGTTCGAGGCGCTGCGGTCTCTGTTGAGCGGCTTCGAGGCTCAGCGT 240
Db 4024 ACCAGTGGTTCGAGGCGCTGCGGTCTCTGTTGAGCGGCTTCGAGGCTCAGCGT 4083

QY 241 CGGCGGACACAGGCTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 300
Db 4084 AACGGACACAGGCTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 4143

QY 301 AAGCGCTTGACGCGCGCGCGCTCCAGCGCGCTCCAGCGCGGTTCGCAAGGCACTGGCC 360
Db 4144 AAGCGCTTCAGGCTTCGAGCGCGCGCTCCAGCGCGGTTCGCAAGGCACTGGCC 4203

QY 361 GCCCGCGAGTCTCCATCGGAGCTCGAGCGGTGAGCGCGCGCGCGCGCGCGCGCG 420
Db 4204 GACCGCGGCTGACGAGCTCCGAGCGGTTCGAGCGGTTCGAGCGCGCGCGCGCGCG 4263

QY 421 CTGGGCGACCGATCGAGCGCGCGCGCTGCTGCGCACCTACGCGCGCGAGAACCG 474
Db 4264 CTGGGCGACCGATCGAGCGCGCGCGCTGCTGCGCACCTACGCGCGCGAGAACCG 4323

QY 475 GAACGCGCGCTGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCG 501
Db 4324 GAACGCGCGCTGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCG 4350

RESULT 15
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match      60.7%; Score 304.2; DB 3; Length 36778;
Best Local Similarity 76.5%; Pred. No. 7.3e-47;
Matches 388; Conservative 0; Mismatches 113; Indels 6; Gaps 1;

QY 1 TCCTCTTCGCTGCTGCGCTGCACATGGCGGTGCGCGCTCGAGCGCGGAGTCTCC 60
Db 5585 TCCTCTTCGCTGCTGCGCTGCACATGGCGGTGCGCGCTCGAGCGCGGAGTCTCC 5644

QY 61 ATGCGCTTCGCGCGCGCTGATGGTATGGGACACGCTCGAGAGTTCGTCGAGTTCCTCG 120
Db 5645 ATGCGCTTCGCGCGCGCTGATGGTATGGGACACGCTCGAGAGTTCGTCGAGTTCAGC 5704

QY 121 CGGCGGCGGCTGGCGCGCGCGCGCTCAAGCGGTTCGCGCGCGGACGCGCGCGCG 180
Db 5705 CGGCGGCGGCTGGCGCGCGCGCGCTCAAGCGGTTCGCGCGCGGACGCGCGCGCG 5764

QY 181 ACCGGCTGGTTCGAGGCGCTCGGCTGCTGTTGAGCGGCTTCGAGGCTCAGCGT 240
Db 3844 TCCTCTTCGCTGCTGCGCTGCACATGGCGGTGCGCGCTCGAGCGCGGAGTCTCC 3903
```

Db 5765 ACCAGCTGGTCCGAGGGCGTCCGGCGTCTCTCTCGTCTGAGGGCTGTCTGACGCCCGCCGC 5824  
QY 241 CGCGGGCACCGGTCTCTGGCGTGTCTCGCGGTCTCGCGGTCACTCCGACGGCGCGTCTG 300  
Db 5825 AACGGACACCGAGTCTCTGGCGTCTCTCGCGGTCTCGCGGTCTGAAACGAGGACGGCGCGAGC 5884  
QY 301 AACGGCTTGACGGCCCCGAAACGGCCCCGTCCACGAGCGGTGATCCGCAAGGCACCTGGCC 360  
Db 5885 AACGSCCTCACGGCTCCGACGCGGCCCTCGACGAGCGCGTCTCTCGCGCGCGTGGCG 5944  
QY 361 GCCGCGGACTGTCCACATCGAACCTCGACGCGGTGGAGGCGCACGGCACCGGGACGACC 420  
Db 5945 GACGCCCGGCTGACGACCTCTCCGACGTGACGTCTCGAGGCACACGGCACGGGCACGCGA 6004  
QY 421 CTGGGCGACCGGATCGAGGCCGAGGGCGTCTGTGGCCACCTACGGCCAGAACCG-----G 474  
Db 6005 CTCGGCGACCGGATCGAGGCGCAGGCCCTGTGATCGCCACCTACGGCCAGGGCGGTGACGAC 6064  
QY 475 GAAACGCGCTGTGGCTCGGGTCTGGTG 501  
Db 6065 GAACAGCGCTGCGCTCGGGTCTGGTG 6091

Search completed: June 17, 2004, 03:21:55  
Job time : 65.6667 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:31:09 ; Search time 293.667 Seconds  
(without alignments)  
7787.123 Million cell updates/sec

Title: US-10-042-665a-3\_COPY\_8434\_8934

Perfect score: 501  
Sequence: 1 tcgtcttcgctggtcgct.....cgctggtcggtcggtcggtg 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2998549 seqs, 2282253817 residues

Total number of hits satisfying chosen parameters: 5997098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	328.2	65.5	18438	15	US-10-156-761-2886
2	328.2	65.5	125746	15	US-10-156-761-15102
3	328.2	65.5	9025608	15	US-10-156-761-1
4	317	63.3	18435	15	US-10-156-761-412
5	317	63.3	100000	15	US-10-156-761-15103
6	316.6	63.2	23238	15	US-10-156-761-415
7	315.4	62.8	5505	15	US-10-156-761-413
8	314.4	62.8	77536	10	US-09-940-316B-1
9	310.2	61.9	10056	15	US-10-156-761-411
10	309.6	61.8	10692	15	US-10-156-761-414
11	309	61.7	10839	15	US-10-156-761-2882
12	309	61.7	86941	17	US-10-461-194-2
13	307.4	61.4	15872	9	US-09-861-289-1
14	307.4	61.4	15872	9	US-09-860-846-1

15	307.4	61.4	15872	10	US-09-988-384B-1	Sequence 1, Appli
16	307.4	61.4	15872	10	US-09-836-821-1	Sequence 1, Appli
17	307.4	61.4	15872	15	US-10-271-889-44	Sequence 44, Appl
18	305.8	61.0	38506	10	US-09-793-708-19	Sequence 19, Appl
19	305.8	61.0	38506	15	US-10-201-365-1	Sequence 1, Appli
20	305.8	61.0	38506	15	US-10-160-539-19	Sequence 19, Appl
21	304.2	60.7	4674	10	US-09-940-316B-26	Sequence 26, Appl
22	304.2	60.7	4725	10	US-09-940-316B-24	Sequence 24, Appl
23	304.2	60.7	4737	10	US-09-940-316B-30	Sequence 30, Appl
24	304.2	60.7	4767	10	US-09-940-316B-28	Sequence 28, Appl
25	304.2	60.7	4818	10	US-09-940-316B-32	Sequence 32, Appl
26	304.2	60.7	13842	9	US-09-861-289-30	Sequence 30, Appl
27	304.2	60.7	13842	9	US-09-860-846-30	Sequence 30, Appl
28	304.2	60.7	13842	10	US-09-988-384B-30	Sequence 30, Appl
29	304.2	60.7	13842	10	US-09-836-821-30	Sequence 30, Appl
30	304.2	60.7	13842	15	US-10-271-889-30	Sequence 30, Appl
31	304.2	60.7	36778	9	US-09-861-289-5	Sequence 5, Appli
32	304.2	60.7	36778	9	US-09-860-846-5	Sequence 5, Appli
33	304.2	60.7	36778	10	US-09-836-821-5	Sequence 5, Appli
34	304.2	60.7	36778	15	US-10-271-889-48	Sequence 48, Appl
35	304.2	60.7	37948	10	US-09-988-384B-5	Sequence 5, Appli
36	303.2	60.5	77536	10	US-09-940-316B-1	Sequence 1, Appli
37	303.2	60.5	64492	13	US-10-378-083-1	Sequence 1, Appli
38	302.6	60.4	6210	15	US-10-212-962-1	Sequence 1, Appli
39	302.6	60.4	85692	17	US-10-461-194-1	Sequence 1, Appli
40	300	59.9	14520	15	US-10-156-761-2885	Sequence 2885, Ap
41	299.4	59.8	11910	15	US-10-156-761-2879	Sequence 2879, Ap
42	299.4	59.8	125746	15	US-10-156-761-15102	Sequence 15102, A
43	299.4	59.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
44	297	59.3	18717	15	US-10-156-761-927	Sequence 927, App
45	297	59.3	30690	17	US-10-204-862A-1	Sequence 1, Appli

## ALIGNMENTS

### RESULT 1

US-10-156-761-2886  
; Sequence 2886, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2886  
; LENGTH: 18438  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(18438)  
US-10-156-761-2886

Query Match 65.5%; Score 328.2; DB 15; Length 18438;  
Best Local Similarity 79.5%; Pred. No. 6.3e-68;  
Matches 403; Conservative 0; Mismatches 98; Indels 6; Gaps 1;

Oy 1 TCGTCTTCGCTGCTGCTGCACATGCGGTGCGAGCCCTGCGAGCGCGGCGAGCTGTCC 60

Db 12457 TCGTGTCCCTGCTGCTGCTGCACATGCGGTGCGAGCCCTGCGAGCGCGGCGAGCTGTCC 12516







```
QY 301 AACGGCTTGAGCGCCCGAAGCGCCCTCCAGCAGCGCGTGTATCCCGAAGGCACTGGCC 360
Db 916 AACGGCTTGAGCGCCCGAAGCGCCCTCCAGCAGCGCGTGTATCCCGAAGGCACTGGCC 975
QY 361 GCCCGCGGACTGTCCACATCGAGCTCGACCGTGGAGGCGGACACCGGACGACGACC 420
Db 976 GCGCGCCGGCTGTCCGCGGAGGAGGTGACCGCTCGAAGCACACGCGCACGCGCACCGC 1035
QY 421 CTGGGCGACCGGATCGAGGCGGAGCGCTGTGGCCACCTACGCGCAGAACCGGAAACG 480
Db 1036 CTCGGCGACCGGATCGAGGCGGAGCGCTGTGGCGACCTATGACAAAGAGCGCGAAACG 1095
QY 481 CCGCTGTGGCTCGGCTCGGTG 501
Db 1096 CCGCTGTGGCTCGGCTCGGTG 1116

RESULT 8
US-09-940-316B-1/c
; Sequence 1, Application US/09940316B
; Publication No. US20030175901A1
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH
; FILE REFERENCE: 30062-20026.11
; CURRENT APPLICATION NUMBER: US/09/940,316B
; PRIOR FILING DATE: 2001-08-27
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275)...(71465)
US-09-940-316B-1

Query Match 62.8%; Score 314.4; DB 10; Length 77536;
Best Local Similarity 76.8%; Pred. No. 7.6e-65;
Matches 384; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 1 TCGTCTTGCTGGTGCAGTGCACATGGCGGTGCAGGCCCTCGAGCGCGCGGAGTGTCTCC 60
Db 42592 TCGTCTTGCTGGTGCAGTGCACATGGCGGTGCAGGCCCTCGAGCGCGCGGAGTGTCTCC 42533
QY 61 ATGGCGCTCGCGCGGCGGTGATGGTGCAGGCGCGGTGCAGGCGTTCGTCCAGTTCCTCG 120
Db 42532 CTGGCGCTCGGTGCAGGCGGTGATGGTGCAGGCGCGGTGCAGGCGTTCGTCCAGTTCCTCC 42473
QY 121 CGGCGAGCGCGGCTGGCCCCCGGACGCGCGTTCGACGCGGTTCGCCGACCGCGCGGACGCG 180
Db 42472 CGGCGAGCGGACTCTCCCGGAGCGGACGCGTTCGACGCGGTTCGCCGACCGCGCGGACGCG 42413
QY 181 ACCGGTGTGTCGAGGCGGTGGGCTGTCTCGTGGTGGAGCGGTGTCCGAGGCTCAGCGT 240
Db 42412 ACCGGTTCGCGGAGGCGGTGGGCTGTCTCGTGGTGGAGCGGTGTCCGAGGCTCAGCGT 42353
QY 241 CGCGGCGCACCAAGTTCCTCGCCGTGGTCCGCGGTTCGCGGCTCGGCGGTCAACTCCGACGCGCGCTCG 300

RESULT 9
US-10-156-761-411
; Sequence 411, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 411
; LENGTH: 10056
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(10056)
US-10-156-761-411

Query Match 61.9%; Score 310.2; DB 15; Length 10056;
Best Local Similarity 76.4%; Pred. No. 1.1e-63;
Matches 381; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 1 TCGTCTTGCTGGTGCAGTGCACATGGCGGTGCAGGCCCTCGAGCGCGCGGAGTGTCTCC 60
Db 622 TCGTCTTGCTGGTGCAGTGCACATGGCGGTGCAGGCCCTCGAGCGCGCGGAGTGTCTCC 681
QY 61 ATGGCGCTCGCGCGGCGGTGATGGTGCAGGCGCGGTGCAGGCGTTCGTCCAGTTCCTCG 120
Db 682 ATGGCGTTCGCGCGGCGGTGATGGTGCAGGCGGTGCAGGCGTTCGTCCAGTTCCTCG 741
QY 121 CGGCGAGCGCGGCTGGCCCCCGGACGCGCGTTCGACGCGGTTCGCCGACCGCGCGGACGCG 180
Db 742 CGGCGAGCGGCTGGCCCCCGGACGCGGTTCGACGCGGTTCGCCGACCGCGCGGACGCG 801
QY 181 ACCGGTGTGTCGAGGCGGTTCGCGGCTGTCTCGTGGTGGAGCGGTGTCCGAGGCTCAGCGT 240
Db 802 ACCGGTGGCGGAGGCGGTTCGCGGCTGTCTCGTGGTGGAGCGGTGTCCGAGGCTCAGCGT 861
QY 241 CGCGGCGCACCAAGTTCCTCGCCGTGGTTCGCGGCTGGCGGTTCGCGGCTCACTCCGACGCGCGCTCG 300
Db 862 AACCGGCGACGAGTTCCTTCGGTGGTTCGCGGCGGAGCGCGGTCACTCCGACGCGCGCTCG 921
```

QY 301 AACGCTTGACGGCCCGGAAACGGCCGCTCCACGACGGCGTGCATCGCAAGCACTGGCC 360  
 DB 922 AACGCTTGACGGCCCGGAAACGGCCCGGCGACGAAACGGGTTCATCGCCAGGCGCTTGCA 981  
 QY 361 GCCGCGGAGCTTCCACATCGGACGCTCGACGGCGGTGGAGGCGCAGCGCACCGGACGCC 420  
 DB 982 TGGGCGGCGCTGTCTCCGCCAGATCGACGCTCGTGGAGCGCACGGCACGGGCGAGCG 1041  
 QY 421 CTGGGCGACCCGATCGAGCGCGAGGGCGCTGTGGCCACCTACGCGCAGAACCGGAAACG 480  
 DB 1042 CTGGGCGACCCGATCGAGCGCGAGGGCGCTCTCGCCACCTACGCGCGGAGCGGACGGAG 1101  
 QY 481 CCGCTGTGCTCGGCTCGG 499  
 DB 1102 GACCGCGCGCTGTGGTTGG 1120

RESULT 10  
 US-10-156-761-414  
 ; Sequence 414, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 414  
 ; LENGTH: 10692  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(10692)  
 US-10-156-761-414

Query Match 61.8%; Score 309.6; DB 15; Length 10692;  
 Best Local Similarity 77.3%; Pred. No. 1.5e-63;  
 Matches 391; Conservative 0; Mismatches 109; Indels 6; Gaps 1;  
 QY 1 TCCTCTTCGCTGCTGCGCTGCACATGCGCGGTGAGCGCCCTGACGCGCGGAGTGTCTCC 60  
 DB 619 TCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 678  
 QY 61 ATGCGCTTCGCGGCGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
 DB 679 CTGGCCGCTGGCGGCGGCTGATCACCAGATGCTGCGCCCGCCCACTTCGTGCGGCTTACG 738  
 QY 121 CGGACAGCGCGGCTGGGCGCCCGGCGGCTGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGG 180  
 DB 739 CGCAGCGCGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 798  
 QY 181 ACCGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
 DB 799 ACCGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 858  
 QY 241 CGCG 300  
 DB 859 AACGG 918  
 QY 301 AACGCTTGACGGCCCGGAAACGGCCGCTCCACGACGGCGTGCATCGCAAGCACTGGCC 360

DB 919 AACGCTTGACGGCCCGGAAACGGCCCGCTCCACGACGGCGTGCATCGCCAGGCGCTTGCC 978  
 QY 361 GCCGCGGAGCTTCCACATCGGACGCTCGACGGCGGTGGAGGCGCAGCGCACCGGACGCC 420  
 DB 979 GCCGCGGAGCTTACACACCGCACGAGTTCGACGGCGGTGGAGGCGCAGCGCACCGGACGCC 1038  
 QY 421 CTGGGCGACCCGATCGAGCGCGAGGGCGCTGTGGCCACCTACGCGCAGAACCG-----G 474  
 DB 1039 CTGGGCGACCCGATCGAGCGCGAGGGCGCTGTCTCGGACCTACGCGCGGCGGCGCTCCCGAC 1098  
 QY 475 GAAAACCGCGCTGTGGCTCGGCTCGGT 500  
 DB 1099 GACAAGCGCGCTGTGGCTCGGT 1124

RESULT 11  
 US-10-156-761-2882  
 ; Sequence 2882, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 2882  
 ; LENGTH: 10839  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(10839)  
 US-10-156-761-2882

Query Match 61.7%; Score 309; DB 15; Length 10839;  
 Best Local Similarity 77.1%; Pred. No. 2.1e-63;  
 Matches 391; Conservative 0; Mismatches 110; Indels 6; Gaps 1;  
 QY 1 TCCTCTTCGCTGCTGCGCTGCACATGCGCGGTGAGCGCCCTGACGCGCGGAGTGTCTCC 60  
 DB 5203 TCCTCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5262  
 QY 61 ATGCGCTTCGCGGCGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
 DB 5263 CTGGCGCTGCTGCGGCGGCTGACCGTATGCGCCACGCGCGGAGCGCTTCGTGCGAGTTCG 5322  
 QY 121 CGGACAGCGCGGCTGGGCGCCCGGCGGCTGCAAGCGCTTCGCGCGGCGGCGGCGGCGGCGG 180  
 DB 5323 CGGACAGCGGCGGCTTCGCGGCGGAGCGCGGCTGCAAGCGCTTCGCGCGGCTTCGCGAGCGG 5382  
 QY 181 ACCGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
 DB 5383 ACCGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
 QY 241 CGCG 360  
 DB 5443 GACGG 5502  
 QY 301 AACGCTTGACGGCCCGGAAACGGCCGCTCCACGACGGCGTGCATCGCAAGCACTGGCC 360  
 DB 5503 AACGCTTGACGGCCCGGAAACGGCGCTCCGTCGACGCGGCTGATCCGCGAAGCACTGGCC 5562



```
QY 361 GCGCGCGACTGTCACATCGGACGTGCGCGGTGAGGCGCGACCGCACCGGAGCGTCC 420
Db 5563 TCGCGCGGGTGTTCGCGCGCGGACGTGCGACATGTTGAGGCGCGACCGGAGCGTCC 5622
QY 421 CTGGGCGACCGCATGAGGCGCGGAGGCGTCTGCGCCACCTTACGCCACAGAACCG-----G 474
Db 5623 CTGGGCGACCGCATGAGGCGCGGAGGCGTCTGCGCGACAGTACGTCAGGACCGTCCGGCG 5682
QY 475 GAAACGCGCTGTGGCTCGGTCGGTG 501
Db 5683 GACCGTCGGCTGTGGCTGGGTTCGGTG 5709

RESULT 12
US-10-461-194-2
; Sequence 2, Application US/10461194
; Publication No. US20040077058A1
; GENERAL INFORMATION:
; APPLICANT: Hutchinson, Richard C.
; APPLICANT: Reid, Ralph C.
; APPLICANT: Hu, Zhihao
; APPLICANT: Rascher, Andreas
; APPLICANT: Schitmer, Andreas
; APPLICANT: McDaniell, Robert
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
; TITLE OF INVENTION: PRO-GEUDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF
; FILE REFERENCE: 300622009700
; CURRENT FILING DATE: US/10/461,194
; PRIOR APPLICATION NUMBER: 2003-06-13
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/389,255
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/393,929
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/395,275
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/415,326
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/420,820
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/433,130
; PRIOR FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 86941
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-461-194-2

Query Match 61.7%; Score 309; DB 17; Length 86941;
Best Local Similarity 76.0%; Pred. No. 1.3e-63;
Matches 381; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTGCGCTGACATGCGCGGTGAGGCGCGTTCGAGCGCGGAGTGTCC 60
Db 31457 TCGTCTTCGCTGCGCTGACATGCGCGGTGAGGCGCGTTCGAGCGCGGAGTGTCC 31516
QY 61 ATGGCGCTCGCGCGCGCGGTGATGGTATGGGACGCTGAGACGCTTCGTGAGTTCCTG 120
Db 31517 CTGGCGCTCGCGCGCGGTGATGGTATGGGACGCTGAGACGCTTCGTGAGTTCCTG 31576
QY 121 CGGACGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 31577 GCGACGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 31636
QY 181 ACCGCTGTGTCGAGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 31637 ACCGCTGTGTCGAGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 31696

Query Match 61.4%; Score 307.4; DB 9; Length 15872;
Best Local Similarity 76.9%; Pred. No. 4.5e-63;
Matches 390; Conservative 0; Mismatches 111; Indels 6; Gaps 1;

QY 1 TCGTCTTCGCTGCGCGCTGACATGCGCGGTGAGGCGCGTTCGAGCGCGGAGTGTCC 60
Db 9017 TCCTCTTCGCTGCGCGCTGACATGCGCGGTGAGGCGCGTTCGAGCGCGGAGTGTCC 9076
QY 61 ATGGCGCTCGCGCGCGCGGTGATGGTATGGGACGCTGAGACGCTTCGTGAGTTCCTG 120
Db 9077 CTGGCGCTGCGCGCGCGGTGATGGTATGGGACGCTGAGACGCTTCGTGAGTTCCTG 9136
QY 121 CGGACGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 9137 CGGACGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9196
QY 181 ACCGCTGTGTCGAGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 9197 ACCGCTGTGTCGAGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9256
QY 241 CGCGGCGCACAGGCTCTCGCGCGGTGCGCGCGGTGCGCGGTGCGCGGTGCGCGGTG 300
Db 9257 GCGGCGCACAGGCTCTCGCGCGGTGCGCGCGGTGCGCGGTGCGCGGTGCGCGGTG 9316
QY 301 AACGCTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 9317 AACGCTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9376
QY 361 GCGCGCGGACTGTCCACATCGGACGCTGACGCGGTGAGGCGCGCGCGCGCGCGCG 420
```

Db	9377	GACCGCGGCTGTCCCGAGAGACGTGACCGGGTCGAGGGCGACCGGACCGGACCGCG 9436
Qy	421	CTGGGCGACCCGATCGAGCGCGAGCGCTGTCTGGCCACTACGGCCAGAACG-----GG 474
Db	9437	CTCGGCGACCCCATCGAGCGCGGGCGCTGTCTGGCGGCTCGGACGGAACGTTCCGGC 9496
Qy	475	GAACCGCGCTGTGCTCGGCTCGGTG 501
Db	9497	GACCAACCGCGCTGTGCTCGGCTCGCTG 9523
RESULT 14		
US-09-860-846-1		
; Sequence 1, Application US/09860846		
; Patent No. US2002016472A1		
; GENERAL INFORMATION:		
; APPLICANT: Sherman, D.H.		
; APPLICANT: Liu, H.		
; APPLICANT: Xue, Y.		
; APPLICANT: Zhao, L.		
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin		
; FILE REFERENCE: 600.438US1		
; CURRENT APPLICATION NUMBER: US/09/860,846		
; CURRENT FILING DATE: 2001-05-18		
; PRIOR APPLICATION NUMBER: 09/105,537		
; PRIOR FILING DATE: 1998-06-26		
; NUMBER OF SEQ ID NOS: 43		
; SEQ ID NO 1		
; LENGTH: 15872		
; TYPE: DNA		
; ORGANISM: Streptomyces venezuelae		
US-09-860-846-1		
Query Match 61.4%; Score 307.4; DB 9; Length 15872;		
Best Local Similarity 76.9%; Pred. No. 4.5e-63;		
Matches 390; Conservative 0; Mismatches 111; Indels 6; Gaps 1;		
Qy	1	TGCTCTTCGCTGTGCTCGCTGCAATGCGCGTGCAGGCCCTGCGAGCGCGCGAGTGTCTCC 60
Db	9017	TCCTCTCGCTGTGCGCTGCAATCTGCGGTACGCGCGCTGCGGCACGCGCGAGTGTCTCC 9076
Qy	61	ATCGCGCTCGCGCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
Db	9077	CTCGCGCTGCGCGCGCGCGGTGGCGGTGATGCGCGATCCGCGCGCGCTTCGTGGAGTTCTCC 9136
Qy	121	CGGACGCGCGCGGTGCGCCCGGCGCTGCAAGCGCTTCGCCGACGCGCGCGGACGCGC 180
Db	9137	CGGCAAGAGGGCTGGCCCGCGCGCGCTGACGCGCGCTGCGAGCGGTTCGAGAGTTCTCC 9196
Qy	181	ACCGGCTGTGTCGAGCGCGCTGCGGCTGCTCTGCTGCGAGCGCGCTGTCGAGGCTCAGCGT 240
Db	9197	ACCGGCTGGGCGCGAGGCGGTGCGCGGTGCTCTGCGAGCGGTTCGAGCGCGCGCGC 9256
Qy	241	CGCGGCGACACAGGTCTCTCGCGGTGCTCGCGGCTGCTCACTCCGACGCGCGGTGCG 300
Db	9257	GCGGGCGACACAGGTCTCTCGCGCTGCTCACTCCGCGCGCGGTGCTGCGAGCGGTGCGCTCC 9316
Qy	301	AACGGCTTGACGCGCGCGCGCGCGCTGCCAGCGCGCTGATCCGACGCGCGCTGCGCC 360
Db	9317	AACGGCTTGACG 9376
Qy	361	GC CGCGCGGACTGTCCATATCGGACGTGCGAGCGCGGTGAGGCGCGACGCGCACCGGAGCAC 420
Db	9377	GACCGCGGCTGTCTCCCGGAGGACGTGGACGCGGTGAGGCGCGACCGGACCGCGCACCGCG 9436
Qy	421	CTGGGCGACCCGATCGAGCGCGCGCGCTGCTGCCACCTACGCGCGAGAAC-----GG 474
Db	9437	CTCGGCGACCCCATCGAGCGCGGGCGCTGCTCGCGCGCTCCGAGCGGAACCGTTCCGGC 9496
Qy	475	GAACCGCGCTGTGCTCGGCTCGGTG 501
Db	9497	GACCAACCGCGCTGTGCTCGGCTCGCTG 9523
Search completed: June 17, 2004, 03:37:02		
Job time : 311.667 secs		

5	102.4	20.4	1303	28	BZ577421	BZ577421 msh2_5402
6	100.8	20.1	662	9	AI514012	AI514012 GH27052.5
C 7	98	19.6	979	28	BZ568491	BZ568491 pacc2-164
8	88.4	17.6	444	12	BI590480	BI590480 RH04107.5
9	88.4	17.6	444	12	BI570498	BI570498 RH02843.5
10	83.8	16.7	392	12	BI569373	BI569373 RH01461.5
C 11	83.8	16.7	766	10	BZ299394	BZ299394 C01118872
12	82.8	16.5	275	13	BQ356242	BQ356242 CW0-GN016
C 13	82.2	16.4	627	12	BM091310	BM091310 1921a10.x
14	82.2	16.4	988	12	BX384927	BX384927 BX384927
15	82.2	16.4	1201	13	BX332425	BX332425 BX332425
16	79.4	15.8	925	23	CNS0091P	AL053013 Drosophil
17	78.6	15.7	747	14	CF869583	CF869583 tric019xn
18	78.6	15.7	801	14	CB899729	CB899729 tric019xn
19	77.6	15.5	1201	13	BX385521	BX385521 BX385521
20	77	15.4	932	29	CNS0072Q	AL066742 Drosophil
C 21	75.4	15.0	925	29	CNS0091P	AL053013 Drosophil
22	74.6	14.7	639	14	CF879070	CF879070 tric019xn
23	73.6	14.7	516	9	AI532314	AI532314 SD03777.5
24	73.4	14.7	378	10	AW502787	AW502787 UI-HE-B80
25	73.4	14.7	918	29	AG043040	AG043040 Pan trogl
26	73	14.6	1308	29	AG133986	AG133986 Pan trogl
C 27	71.2	14.2	1242	12	BM911414	BM911414 AGENCOURT
28	70.8	14.1	935	29	CNS006XK	AL066051 Drosophil
29	70.2	14.0	976	29	AG041117	AG041117 Pan trogl
30	70	14.0	982	13	BX415111	BX415111 BX415111
31	69.6	13.9	1233	13	AG073513	AG073513 Pan trogl
C 32	69.2	13.8	935	29	CNS006XK	AL066051 Drosophil
C 33	68.8	13.7	932	29	CNS0072Q	AL066742 Drosophil
C 34	68.4	13.7	1370	29	AG072551	AG072551 Pan trogl
35	67.8	13.5	1100	29	CNS016KD	AL106855 Drosophil
C 36	67.6	13.5	851	14	CB908312	CB908312 tric084xp
37	67.2	13.4	722	14	CB898379	CB898379 tric014xn
C 38	67.2	13.4	924	13	BX442207	BX442207 BX442207
39	67	13.4	1075	29	AG073983	AG073983 Pan trogl
C 40	66.8	13.3	982	13	BX415111	BX415111 BX415111
41	66.2	13.2	667	14	CF868305	CF868305 tric014xn
C 42	65.6	13.1	799	29	AG043038	AG043038 Pan trogl
43	65.2	13.0	695	9	AL507566	AL507566 AGENCOURT
44	65.2	13.0	1136	13	BQ943816	BQ943816 AL507566
45	64.8	12.9	699	29	CC731567	CC731567 OGUKKI9TV

## ALIGNMENTS

```

FEATURES             1..508
     seq_1            linear      GSS 17-DEC-2002
     locus            B2561977
     definition        pac82-164_3530.x1 pac82-164 Pseudomonas aeruginosa genomic clone
     accession         B2561977_3530, genomic survey sequence.
     version           B2561977.1 GI:27183081
     keywords          GSS.
     organism          Pseudomonas aeruginosa
     source            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
                       Pseudomonadaceae; Pseudomonas.
     reference         1 (bases 1 to 753)
                       Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
                       Burns,J.L., Kaul,R. and Olsen,M.V.
                       Whole-Genome-Sequence variation among multiple isolates of
                       Pseudomonas aeruginosa library
     title             J. Bacteriol. (2002) In press
     journal           Contact: Chris K. Raymond
     comment           Genome Center
                       University of Washington
                       Box 352145, Seattle, WA 98105-2145, USA
                       Tel.: 2062216954
                       Fax: 2066857244
                       Email: craymond@u.washington.edu
                       Class: shotgun.
                       Location/Qualifiers

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM nucleic - nucleic search. using sw model

Run on: June 16, 2004, 23:18:23 ; Search time 1960 Seconds  
(without alignments) 7633.139 Million cell updates/sec

Title: US-10-042-665A-3 COPY 8434 8934

perfect score: 501  
Sequence: 1 tcgtctcgctggtcgcgct.....cgctgtggtcgggtcggtg 501

Scoring table: IDENTITY\_NUC  
Gapop 10.0 : Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:	55026578
--	----------

Minimum DB seq length: 0

Maximum DB seq	length: 2000000000
----------------	--------------------

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

```

1:  em_estba:*
2:  em_esthum:*
3:  em_estin:*
4:  em_estmu:*
5:  em_estov:*
6:  em_estpl:*
7:  em_estro:*
8:  em_htc:*
9:  gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun
16: em_estom:*
17: em_gss_hum
18: em_gss_inv
19: em_gss_pla
20: em_gss_vr
21: em_gss_fun
22: em_gss_mana
23: em_gss_mu
24: em_gss_pro
25: em_gss_ro
26: em_gss_ph
27: em_gss_vr
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Query			DB	ID	Description
		Score	Match	Length			
	1	141.6	28.3	753	28	BZ561977	pac2c.164
	2	116	23.2	871	28	BZ566693	pac2c.164
C	2	104.2	20.8	625	12	EM091001	ig21a10.Y
	3	103.6	20.7	865	28	BZ566683	pac2c.164
C	4	103.6	20.7	865	28	BZ566683	pac2c.164



obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@imgate.wustl.edu)  
Possible reversed clone: similarity on wrong strand  
High quality sequence stop: 485.

#### FEATURES

Location/Qualifiers  
1..625  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="Fetal Pancreas (4 Pooled Donors, 18 - 20 weeks, Stragene #738023)"  
/dev\_stage="Fetal Pancreas"  
/clone\_lib="Human Fetal Pancreas 1B"  
/note="Vector: pBluescript SK(-); Site 1: NotI; Site 2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size -1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916, Fax: 314-747-2692."

#### ORIGIN

Query Match 20.8%; Score 104.2; DB 12; Length 625;  
Best Local Similarity 52.3%; Pred. No. 0.00014;  
Matches 260; Conservative 0; Mismatches 228; Indels 9; Gaps 1;

```
Qy 1 TCGTCTTCGCTGTCGCGCTGCACATGCGCGTGTGAGCCCTGACGCGGCGGAGTGTCC 60
Db 604 TCTTCCAGCTGTATGGCCCTGCAGACGCTTACCAGGCCATCCACGCGGCGGCGCT 545
Qy 61 ATGCGCTGCGCGCGCGGTGATGTGATGGCAGCGTCTGAGAGCGTTCGTGAGTTCG 120
Db 544 GCCGCCATCTGTTGGGGGATCAAGCTCTCTGTAAGCCCAACACTCGTGCAGTTCG 485
Qy 121 CGGCAGCGCGGCGTGGCCCCCGACGCGCGCTGCAAGCGTTCGCCGACGCGCGAGCGC 180
Db 484 AGGCTGGGATGCTCAGCCCCGAGGCGACCTTCAAGGCTTCGACACAGCGGGAATGG 425
Qy 181 ACCGGCTGTCGAGGGCGTGGGCTGCTCTGTTGAGCGGTGTCGAGGCTCAGGCT 240
Db 424 TACTGCGGCTCGAGGGTGTGTGCGCGTCTCTGTGACCAAGAGTCTCCCTGCGCGCGG 365
Qy 241 CGCGGACACAGTCTCGCGTGTGCGGCGTGTGCGGCGTCACTCCGACGCGCGCTCG 300
Db 364 GTGTAGGCCACATCTCTGAACGCGGACCAATACAGATGGCTTCAAGGA-----G 314
Qy 301 AACGGCTTGAAGGCGCCCGGAGCGCCCGTCCAGCAGCGCGTGTATCCGCAAGGCACTGGCC 360
Db 313 CAAGCGTGACTTCTCCCTCAGGGGATATCCAGGAGCAGCTCATCCGCTGTTGTACCAG 254
Qy 361 GCGCGCGAGTGTGCATATCGAGCTGACGCGGTGAGGCGGACGCGACCGGAGCAGCC 420
Db 253 TCGCGCGAGTGGCCCTTGAGTTCATTGATATACATCGAAGCCCAACGCGACAGCCCAAG 194
Qy 421 CTGGCGACCCGATCGAGCGGAGCGCTGTGCGCCACCTACGCGCAGAACCGGGAACG 480
Db 193 GTGGCGGACCCCGAGAGCTGAATGGCATCACCAGCGCTGTGCGCCACCCCGCAGGAG 134
Qy 481 CGCTGTGCTCGGCTC 497
Db 133 CGCTGTCTATCGGCTC 117
```

#### RESULT 4

BZ566683  
LOCUS BZ566683 865 bp DNA linear GSS 17-DEC-2002  
DEFINITION pacs2-164\_6520.x1 pacs2-164 Pseudomonas aeruginosa genomic clone  
pacs2-164\_6520, genomic survey sequence.  
ACCESSION BZ566683  
VERSION BZ566683.1 GI:27196700  
KEYWORDS GSS.

#### SOURCE

ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

#### REFERENCE

AUTHORS 1 (bases 1 to 865)  
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.

#### TITLE

Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library

#### JOURNAL

COMMENT J. Bacteriol. (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

#### FEATURES

source  
1..865  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pacs2-164\_6520"  
/clone\_lib="pacs2-164"  
/note="clinical isolate 2-164 Whole genomic shotgun  
library."

#### ORIGIN

Query Match 20.7%; Score 103.6; DB 28; Length 865;  
Best Local Similarity 63.0%; Pred. No. 0.00017;  
Matches 160; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy 248 ACCAGGTCTTCGCGGTGTGTCGCGGCTCGGCGTCACTCCGAGCGGCGCTCGAACGCT 307
Db 92 ACCGGGTATCGCGGCTTCTGCTGGGCTCGGCGGTCAACCATGACGCGCCCTGCGGGGGC 151
Qy 308 TGACGGCCCCGAAAGCGCCGCTCCAGCAGCGCGTGCATCCGCAAGGCACTGGCGCGCGCG 367
Db 152 TCAGGTATCCCAAGCAGCGCCGCGGAGGCGGTTGATTCGGAAGCGCTGGCCACGCCA 211
Qy 368 GACTGTCCATATCGAGCGTGCAGCGGTGAGGCGCACCGGACGAGACCTTGGGCG 427
Db 212 GCTGTGCGCGCGCGGTGCGGCTATGTCGAGGCGCCATGCGGACCGGCACAGTGTCTCGCG 271
Qy 428 ACCGATCGAGCGCGGCGCTGTGCGCACCTACGGCCAGNACCGGGAAGCGCGCTGT 487
Db 272 ATCCGATCGAGTTGCGAGCCCTGCGCCACGCGCTATCGAGGTACTGAGCGCGTCCCTTGG 331
Qy 488 GCCTCGGCTCGGTG 501
Db 332 CGTGGCATCGGTG 345
```

#### RESULT 5

BZ577421  
LOCUS BZ577421 1303 bp DNA linear GSS 17-DEC-2002  
DEFINITION msh2\_5402.x1 msh Pseudomonas aeruginosa genomic clone msh2\_5402,  
genomic survey sequence.

#### ACCESSION

VERSION BZ577421  
GI:27212482

#### KEYWORDS

GSS.

#### SOURCE

ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

#### REFERENCE

AUTHORS 1 (bases 1 to 1303)  
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol. (2002) In press

## COMMENT

Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun

## FEATURES

source  
 1..1303  
 /location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone="msh2\_5402"  
 /clone\_lib="msh"  
 /note="Environmental isolate. Whole genomic shotgun library."

## ORIGIN

Query Match 20.4%; Score 102.4; DB 28; Length 1303;  
 Best Local Similarity 62.3%; Pred. No. 0.00025;  
 Matches 157; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTGCTGCACATGGCGGTGCAGGCCCTGCAGCGCGCGAGTGTCTCC 60  
 Db 362 TCGTCTTCGCTGCTGCACATGGCGGTGCAGGCCCTGCAGCGCGCGAGTGTCTCC 421

QY 61 ATGGCGCTCGCGCGCGCGGTGATGGTGGGACCGTGCAGACGTTGTCGAGTTCCTCG 120  
 Db 422 CTGNCCTGCGCGCGCGGTGACCTGATCTTTCGCCCTATGTTCTGCTCTGTCGACG 481

QY 121 CGCAGCGCGGCTGCCCCCGAGCGCTGCAAGCGTTGCCGACGCGCGGACGCGC 180  
 Db 482 AAGTTGCGGCGCTGCACGCGATGGCGGTGCAAGACCTTCACGCGCGCGGACGCGC 541

QY 181 ACCGCTGTCGCGCGCGCGTCTCTGTTGGAGCGGTGTCGCGAGGCTTCAGCGT 240  
 Db 542 TATGGCTGNGAGAGGCTGCTGTTGCTTCAAGCGCTGCGCGAGCGCGGCGT 601

QY 241 CGCGGGCACCG 252  
 Db 602 GACCGTGATCCG 613

## RESULT 6

AI514012 662 bp mRNA linear EST 23-APR-2001  
 LOCUS GH27052.5prime GH Drosophila melanogaster head pOT2 Drosophila  
 DEFINITION melanogaster cDNA clone GH27052 5 similar to BcDNA:GH07626;  
 Phn0003523 located on: 2L 23D1-23D1; 04/10/2001, mRNA sequence.  
 AI514012  
 AI514012.1 GI:4418074  
 EST.

ORGANISM Drosophila melanogaster (fruit fly)  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 662)  
 Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
 Lewis, S. and Rubin, G.M.  
 BDGP/HMI Drosophila EST Project  
 Unpublished (2001)  
 Other ESTs: GH27052.3prime  
 Contact: Stapleton, M.  
 BDGP

## REFERENCE

AUTHORS Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6796  
 Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
 hit genomic AE003581: arm:2L [2747932,3054367]  
 estimated-cyto:23B3-23D3: 04/10/2001

Plate: GH.270 row: E column: 4  
 High quality sequence stop: 542  
 POLYA=No.

## FEATURES

source  
 1..662  
 /location/Qualifiers  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7227"  
 /clone="GH27052"  
 /sex="male and female"  
 /dev\_stage="adult"  
 /lab\_host="DHS - alpha"  
 /clone\_lib="GH Drosophila melanogaster head pOT2"  
 /note="Organ: head; Vector: pOT2; Site 1: EcoRI; Site 2:  
 XhoI; Sized fractionated cDNAs were directly ligated into  
 pOT2. Plasmid cDNA library."

## ORIGIN

Query Match 20.1%; Score 100.8; DB 9; Length 662;  
 Best Local Similarity 51.8%; Pred. No. 0.00038;  
 Matches 259; Conservative 0; Mismatches 232; Indels 9; Gaps 1;

QY 1 TCGTCTTCGCTGCTGCACATGGCGGTGCAGGCCCTGCAGCGCGCGAGTGTCTCC 60  
 Db 6 TCCAGTTCTCTTACGCCCTTGGAACAGGCTTTCGCGATATGCGGAGAAAGTCCGAC 65

QY 61 ATGGCGCTCGCGCGCGCGTGTGATGGGACCGTGCAGACGTTTCGAGTTCCTCG 120  
 Db 66 AAGCCCTGCTGCTGGAGCTGCATCTCAAGCCACCATGTCGCTGCAGTTCAAG 125

QY 121 CGCAGCGCGGCTGCCCCCGACGCGCTGCAAGCGTTGCCGACGCGCGGACGCGC 180  
 Db 126 CGACTGAATATGTTGAGCCCGGACGCGCTGCAAGCGCTTCGATGAGTCTGGCAATGA 185

QY 181 ACGGCTGTCGAGGCGCTCGGGCTGCTGCTGAGCGGCTGCCAGGCTCAGCGT 240  
 Db 186 TAGTCCGTTCCGATGGATGTTGGTGTGCTGCTCTCAGCGCACCTCTCGAGCAGCGT 245

QY 241 CGCGGGCACCGTCTCTCGCGGTGTCGCGGCTCGGCGTCAACTCCGACGCGCGTCCG 300  
 Db 246 GTGATGCTCCATCTCAATGTCGCGCACCAATACGATGTTTCAAGGA-----G 296

QY 301 AACGGTTGAAGCGCCCCGAAACGCGCTGCCAGCGCGGTGATCGCAAGGACTGGCC 360  
 Db 297 CAGGCGATCACATACCTATTGGCAAGATGCAAAATCGCTGATCGCGAGACTACGAG 356

QY 361 GCGCGGAGTGTCCACATCGGACGTCGACGCGGTGAGCGCGCACCGCACCGGACGCC 420  
 Db 357 GAGATTGGTCTTAACCCCGCGGATGTTTACGTGGAGSCACACGTTACCGNACCAAG 416

QY 421 CTGGCGACCCGATCGAGCGCGGCGCTGCTGCCACCTTACGCGCCAGAACCGGAAACG 480  
 Db 417 GTGGCGATCCCCAGGAGGTGAACCTCTATCACTGACTTCTTCTGCAAGGACCGTACGACC 476

QY 481 CCGCTGTGCTCGGTCGCT 500  
 Db 477 CCTGTGTGATCGGATCGGT 496

## RESULT 7

BZ568491/c 979 bp DNA linear GSS 17-DEC-2002  
 LOCUS BZ568491  
 DEFINITION pac82-164\_7558.y2 pac82-164 Pseudomonas aeruginosa genomic clone  
 pac82-164\_7558, genomic survey sequence.  
 ACCSSION BZ568491  
 VERSION BZ568491.1 GI:27201527  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 1 (bases 1 to 979)  
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,



Burns, J.L., Kaul, R. and Olsen, M.V.  
 Whole-Genome-Sequence variation among multiple isolates of  
*Pseudomonas aeruginosa* library  
 J. Bacteriol. (2002) In press  
 CONTACT: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2068857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

**FEATURES**  
 source  
 1..979  
 Location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="pacs2-164 7558"  
 /clone\_lib="pacs2-164"  
 /notes="Clinical isolate 2-164 Whole genomic shotgun library."

**ORIGIN**  
 Query Match 19.6%; Score 98; DB 28; Length 979;  
 Best Local Similarity 58.1%; Pred. No. 0.00093;  
 Matches 173; Conservative 0; Mismatches 125; Indels 0; Gaps 0;  
 QY 1 TCGTCTTCGTGTCGCGCTGCATCGCGGTGCGAGCCCTGCGAGCGCGGAGTGTCTCC 60  
 Db |||||  
 QY 389 TCGTCTTCGTGTCGCGCTGCATCGCGGTGCGAGCCCTGCGAGCGCGGAGTGTCTCC 330  
 Db |||||  
 QY 61 ATGGCGCTCCGCGCGCGGTGATGTCATGCGCAGCGTTCGAGAGTTCGTGCGAGTTCCTCG 120  
 Db |||||  
 QY 329 ATTGCTATCGTGGCGGCGGTGAAGTATCGATCTCGCGCAATCTTCCATCCATGGGC 270  
 Db |||||  
 QY 121 CGGCGAGCGCGGCTGGCGCCCGCGCGCTGCAAGCGGTTCGCCGACGGCGCGGACGGC 180  
 Db |||||  
 QY 269 GATCCCGTGGCTATCGGCGGACGGCACCTGCAAGACCTTCGACGATCGTGGCGATGGC 210  
 Db |||||  
 QY 181 ACCGCGTGTCCGAGGCGTGGCGCTGCTCTCGTGGAGCGGCTGTCCGAGGCTCAGCGT 240  
 Db |||||  
 QY 209 TACGCGCGCGGAGGCTTGGCGGCTGTGTGATCTCAGCGCGGCGGACCGGCGCGGCGC 150  
 Db |||||  
 QY 241 CGCGGCGACCGAGTCCGCGCTGTCCGCGGCTCGCGGCTCACTCCGACGCGCGCT 298  
 Db |||||  
 QY 149 GACGCGACCGGCTATCGCGGCTTCTGTGGCTCGGCGGTCAACCATGACGGCGCT 92  
 Db |||||

**RESULT 8**  
 BI590480 444 bp mRNA linear EST 06-SEP-2001  
 LOCUS RH04107.5prime RH Drosophila melanogaster normalized Head pFlc-1  
 DEFINITION Drosophila melanogaster cDNA clone RH04107 5 similar to  
 BcDNA:GH07626; FBan0003523 GO:[fatty-acid synthase (GO:0004312)]  
 located on: 2L 23D1-23D1;: 08/16/2001, mRNA sequence.  
 BI590480  
 BI590480.1 GI:15481902  
 EST.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 444)  
 STAPLETON, M., BROKSTEIN, P., HONG, L., TYLER, D., BERMAN, B.,  
 CARLSON, J., CHAMPE, M., CHAVEZ, C., DORSETT, V., FARFAN, D., FRISSE, E.,  
 GEORGE, R., GONZALEZ, M., GUARIN, H., HARRIS, N., LI, P., LIAO, G.,  
 MISRA, S., MUNGALL, C.J., NUNOO, J., PACLEB, J., PARAGAS, V., PARK, S.,  
 PHOUANAVONG, S., WAN, K., YU, C., LEWIS, S.E., CELINKER, S. and  
 RUBIN, G.M.  
 TITLE BDGP/HMI RH Drosophila EST Project  
 JOURNAL Unpublished (2001)

CONTACT: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: http://www.fruitfly.org/EST. est@fruitfly.berkeley.edu  
 hit genomic AE003581: arm:2L [2747932,3054367]  
 estimated-cyto:23B3-23D3: 08/16/2001  
 Plate: RH.41 row: A column: 7  
 high quality sequence stop: 423.  
 Location/Qualifiers  
 1..444  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7227"  
 /clone="RH04107"  
 /sex="male and female"  
 /dev\_stage="Adult"  
 /lab\_host="DH5-alpha Tona"  
 /clone\_lib="RH Drosophila melanogaster normalized Head pFlc-1"  
 /notes="Organ: head; Vector: pFlc1; site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

**FEATURES**  
 source  
 1..444  
 Location/Qualifiers  
 103 ACGTTCGTGAGTTCCTCGGCGAGCGGCTGCCCCCGACGCGCGCTGCAAGCGCTTC 162  
 Db |||||  
 QY 7 ATGTGCTGCTGAGTTCAAGCGACTGATATGTTGAGCCCGGACGGCAGCTGCAAGGCTTC 66  
 Db |||||  
 QY 163 GCCACGCGCGGACGCGCTGTCGAGGCGGTGCGGCTGCTCTCTGTGAGCGG 222  
 Db |||||  
 QY 67 GATGAGTATGCAATGATGATGTCGCTTCCATCTCAATGTGCGCACCACATACGATGGT 126  
 Db |||||  
 QY 223 CTGTCCGAGGTCACGCGTCGCGGCGACGAGTCTCTCGCGGTGTCGCGGTCGCGCGTC 282  
 Db |||||  
 QY 127 ACCTCTGCGCGAGCGGTGTGATGTCCTTCCATCTCAATGTGCGCACCACATACGATGGT 186  
 Db |||||  
 QY 283 AACTTCGACGCGCGTGAACGCGTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 342  
 Db |||||  
 QY 187 TTCAAGGA-----GCGGCGCATCATACCTTATTTGCGAAGATGCAAAATCGCCTG 237  
 Db |||||  
 QY 343 ATCCGCGAGGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 402  
 Db |||||  
 QY 238 ATCCGCGAGGCACTGCGAGGAGATGCTTAAACCCCGCGATGCTTACGTGAGGCA 297  
 Db |||||  
 QY 403 CACGCGACCGGAGACCCCTGCGCGACCCCGATCGAGGCGCGCGCGCGCGCGCGCGCG 462  
 Db |||||  
 QY 298 CACGCTACCGGACCAAGTGGCGGATCCCGCAGGAGTGAACCTATCACTGACTTCTTC 357  
 Db |||||  
 QY 463 GCGCAGAACCGGCGAAACCGCGCTGTGGCTCGGCTCGGT 500  
 Db |||||  
 QY 358 TCGAAGGACCGTACGACCCCTCTGTTGATCGGATCGGT 395  
 Db |||||

**RESULT 9**  
 BI570498 502 bp mRNA linear EST 06-SEP-2001  
 LOCUS RH02843.5prime RH Drosophila melanogaster normalized Head pFlc-1  
 DEFINITION Drosophila melanogaster cDNA clone RH02843 5 similar to  
 BcDNA:GH07626; FBan0003523 GO:[fatty-acid synthase (GO:0004312)]  
 located on: 2L 23D1-23D1;: 07/26/2001, mRNA sequence.  
 BI570498  
 BI570498.1 GI:15461920  
 EST.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;



```

QY 403 CACGGCACCGGAGACACCTCGGGGACCCGATCGAGGCCGAGGCGCTGCTGGCCACCTAC 462
Db 298 CACGGTACCGGAACCAAGTGGCGGATCCCGAGGAGGTGAATCTTATCACTGACTTCTTC 357
QY 463 GCGCAGACCGGGAACCGCGCTGTGGCTCGGTC 497
Db 358 TCGAAAGACCGTACGACCCCTCTGTGTATCGGATC 392

RESULT 11
LOCUS BE299394/c 766 bp mRNA linear EST 20-JUL-2000
DEFINITION 601118872F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3028841 5',
ACCESSION BE299394
VERSION BE299394.1 GI:9183142
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 766)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cyabps-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM87 row: o column: 18
High quality sequence start: 6
High quality sequence stop: 674.
FEATURES
    source
        1. .766
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:3028841"
            /tissue_type="rhabdomyosarcoma"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH MGC 17"
            /note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI;
            Site 2: XhoI; cDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCAGAG(G). Size-selected >500bp
            for average insert size 1.8kb. Library constructed by
            Ling Hong in the laboratory of Gerald M. Rubin (University
            of California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
```

## ORIGIN

```

Query Match 16.7%; Score 83.8; DB 10; Length 766;
Best Local Similarity 59.4%; Pred. NO. 0.066;
Matches 142; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTGCTCGCGCTGATGTGATGGGACCGTTCGAGACCTTCGTCAGTTCTCG 120
Db 395 TCCTCCAGCCTGATGGCCCTCGAAGCCCTACCGAGCCATCCACAGCGGCGAGTGCCT 336

QY 61 ATGGCGCTCGCGGGCGGCTGATGTGATGGGACCGTTCGAGACCTTCGTCAGTTCTCG 120
Db 335 GCCGCATCTGTGGGGGGGATCAACGCTCTGTGTGAGCCCAACACCTCCGTCAGTTCTG 276

QY 121 CGGCAAGCGCGGCTCGCCCCCGACGCGCTCGAAGCGGTTCCGCGACGCGCGGACGCG 180
Db 275 AGGCTGGGGATGCTCAGGCCCGGAGGCGCTCGAAGGCGCTTCGACACAGCGGGGATGG 216

QY 181 ACCGGCTGTGCGAGGCGGTGGGCTGTCTTCGTGGAGCGGCTGTCCGAGCTCAGCG 239
```

```

Db 215 TACTGCCCTCGAGGGTGTGTGGCCCTCTCTGCTGACCAAGAGTCCCTGGCCCGGCG 157

RESULT 12
LOCUS BQ366242 275 bp mRNA linear EST 21-MAY-2002
DEFINITION CM0-GN0162-121000-624-d04 GN0162 Homo sapiens cDNA, mRNA sequence.
ACCESSION BQ366242
VERSION BQ366242.1 GI:21041754
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 275)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0&t2=CM0-GN0162-
121000-624-d04&t3=2000-10-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 275.
FEATURES
    source
        1. .275
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /dev_stage="Adult"
            /clone_lib="GN0162"
            /note="Organ: placenta normal; Vector: puc18; Site 1:
            SmaI; Site 2: SmaI; A mini-library was made by cloning
            products derived from ORESTES PCR (U.S. Letters Patent
            application No. 196.716 - Ludwig Institute for Cancer
            Research) profiles into the puc 18 vector. Reverse
            transcription of tissue mRNA and cDNA amplification were
            performed under low stringency conditions."
```

## ORIGIN

```

Query Match 16.5%; Score 82.8; DB 13; Length 275;
Best Local Similarity 60.0%; Pred. NO. 0.08;
Matches 138; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 10 CTGTCGCGCTGCACATGGCGGTGCAGGCCCTTCGAGCGCGCGAGTGTCTCATGGCGCTC 69
Db 6 CTGATGGCCCTGCAGAACGCTTACCAGGCCATCCACAGCGGCGAGTGCCTCCGCCCATC 65

QY 70 GCGGCGCGCTGATGTGATGGGACCGTTCGAGAGCTTCGTCGAGTTTCGCGGCGAGCGC 129
Db 66 GTGGGGGGCATCAATGTCTCTGTAAGCCCAACACCTCCGTGCACTTCTTGAGGCTGGGG 125

QY 130 GGGCTGGCCCCCGCGCGCTCGAAGCGGTTTCGCCGACGCGCGGACGCGACCGCGCTGG 189
Db 126 ATGCTCAGCCCCGAGGGGCACTTCGAAGGCTTCGACACAGCGGGGAATGGGTACTGCCGC 185
```

QY 190 TCCGAGGGCTCGGCTGCTCTGGTGGAGCGGCTGTCGAGGCTCAGCG 239  
 Db 186 TCGAGGGGTGGTGGCGGCTCTGCTGACCAAGAAGTCCCTGGCCCGCG 235

RESULT 13  
 BM091310  
 LOCUS 1921a10.x1 Human Fetal Pancreas 1B mRNA linear EST 20-NOV-2001  
 DEFINITION TR:Q16702 Q16702 FATTY ACID SYNTHASE 1, mRNA sequence.

ACCESSION  
 BM091310  
 VERSION 1  
 KEYWORDS  
 SOURCE

ORGANISM Homo sapiens (human)

REFERENCE  
 AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Konko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y. and Bowers, Y.

TITLE Endocrine Pancreas Consortium

JOURNAL Unpublished (2000)

COMMENT Other ESTs: 1921a10.y1

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@biohp.harvard.edu  
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Dr. Hiroshi Inoue  
 (hinoue@imgate.wustl.edu)  
 Possible reversed clone: similarity on wrong strand  
 High quality sequence stop: 480.

FEATURES  
 source

1..627  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /tissue\_type="Fetal Pancreas (4 Pooled Donors, 18 - 20 weeks, Stratagene #738023)"  
 /dev\_stage="Fetal Pancreas"  
 /clone\_lib="Human Fetal Pancreas 1B"  
 /notes="Vector: pBluescript SK(-); Site 1: NotI; Site 2: XhoI; cDNA made by oligo-dT priming. Size selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave, St. Louis, MO 63110 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916, Fax: 314-747-2692."

ORIGIN

Query Match 16.4%; Score 82.2; DB 12; Length 627;  
 Best Local Similarity 59.0%; Pred. No. 0.1;  
 Matches 141; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTCGCGCTGCACATGCGCGTGCAGGCCCTGCAGCGCGCGAGTGTCC 60  
 Db 295 TCCTCCAGCCTGATGGCCCTGCAGAGCCCTACAGGCCCATCCAGCGCGCGAGTGCCT 354

QY 61 ATGGCGCTCCGGCGCGCTGTGATGGTGTGATGGGACGGTTCGAGAGGTTCTCG 120  
 Db 355 GCGGCCATCTGGGGGGCATCAACGTCCTGCTGAAGGCCCAACACCTCCGTGCAGTTCTTG 414

QY 121 CGGCAGCGCGGGCTGGCCCGCGGCTGCAAGCGGTTGCGCCGACGGCGCGAGCGC 180  
 Db 415 AGGCTGGGATGCTCAGCCCCGAGGGCACCTGCAAGGCTTCGACACAGCGGGGAATGGG 474  
 QY 181 ACCGGCTGGTCCGAGGGCGTGGGCTGCTCTGTTGTGAGGGCTGTCCGAGGCTCAGCG 239  
 Db 475 TACTGCCGCTCGGAGGGTGTGGTGGCGCTGCTGTGACCAAGAAGTCCCTGGCGCGCG 533

RESULT 14  
 BX384927  
 LOCUS

DEFINITION BX384927 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 Homo sapiens cDNA clone CS0DL004YA22 5-PRIME, mRNA sequence.

ACCESSION  
 BX384927  
 VERSION 1  
 KEYWORDS  
 SOURCE

ORGANISM Homo sapiens (human)

REFERENCE  
 AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 8574.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DL004B11Q1&cluster=8574.r. Contact :  
 Feng liang Email: fliang@lifetech.com URL: Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID: CS0DL004B11Q1.

FEATURES  
 source

1..988  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL004YA22"  
 /cell\_type="B CELLS (RAMOS CELL LINE)"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 16.4%; Score 82.2; DB 13; Length 988;  
 Best Local Similarity 59.0%; Pred. No. 0.11;  
 Matches 141; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTCGCGCTGCACATGGCGGTGCGAGGCCCTGCAGCGCGCGAGTGTCC 60  
 Db 652 TCCTCCAGCCTGATGGCCCTGCAGAGCCCTACAGGCCCATCCAGCGCGCGAGTGCCT 711

QY 61 ATGGCGCTCCGGCGCGCGCTGATGGTGTGATGGGACGGTTCGAGAGGTTCTCG 120  
 Db 712 GCGGCCATCTGGGGGGCATCAATGCTGCTGTAAGCCCAACACCTCCGTGCAGTTCTTG 771

QY 121 CGGCAGCGCGGGCTGGCCCCCGACGGCGCTTCGAGGCGTTCCGCCACGGCGCGGACGGC 180  
 Db 772 AGGCTGGGATGTTCAGCCCCGAGGGCACCTTCGAGCCCTTCACACAGCGGGGAATGGG 831

QY 181 ACCGGCTGTCCGAGGGCGTCCGGGCTGCTCTGTTGTGAGGGTGTCCGAGGCTCAGCG 239  
 Db 832 TACTGCCGCTCGAGGGTGTGGTGGCGCTCTGCTGACCAAGAAGTCCCTGGCGCGCG 890

Thu Jun 17 08:25:05 2004

```

RESULT 15
BX332425      1201 bp      mRNA      linear      EST 02-MAY-2003
LOCUS      BX332425 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION      cDNA clone CSODC012YI01 5-PRIME, mRNA sequence.
ACCESSION      BX332425
VERSION      BX332425.1 GI:30335143
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 1201)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 8574.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODC012AE01QP1&cluster=8574.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CSODC012AE01QP1.
FEATURES
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODC012YI01"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match      16.4%; Score 82.2; DB 13; Length 1201;
Best Local Similarity 59.0%; Pred.No. 0.11;
Matches 141; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY      1      TCGTCTTCGTGCGCGCTGCACATGCGGTGCGAGCCCTGCAGCGCGCGAGTGTCTCC 60
Db      660      TCCTCCAGCCTGATGGCCCTGCAGACGCCTACAGGCCATCCACAGGGGCACTGCCCT 719
QY      61      ATGGCGCTTCGCGGCGCGGTGATGTTGATGGGCACGGTCGAGACGTTGCTGAGTTCTCG 120
Db      720      GCCCCATCGTGGGGGCATCATGTCTGCTGAAGCCCAACACCTCCGTGAGTTCTTG 779
QY      121      CGGAGCGCGGGGTGGCCCCCGAGCGCGCTGCAGAGGGTTGCGCGAGCGCGGACGCGC 180
Db      780      AGGCTGGGGATGCTCAGCCCCGAGGGCACCTGCAAGCCCTTCGACACAGCGGGGAATGGG 839
QY      181      ACCGGCTGTCGAGGGGCGTCGGGGTCTGCTGTGGTGGAGCGGCTGTCAGGCTCAGCG 239
Db      840      TACTGCCGCTCGAGGGGTGTGTGGCCGCTCTGCTGACCAAGAGTCTCTGGCCGGCG 898

```

Search completed: June 17, 2004, 03:18:50  
Job time : 1966 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 22:09:53 ; Search time 2276.33 Seconds  
(without alignments)  
9539.403 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543  
Perfect score: 501  
Sequence: 1 gcggccccgacgcgagca.....ccctcggcgacgagctga 501

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	53784	1	AMM223012
2	501	100.0	53789	6	A69720
3	501	100.0	90445	1	AF040570
4	275.6	55.0	22131	1	AF506522
5	268.8	53.7	6567	1	AF521897
6	164.4	32.8	32870	1	AF007101
7	158.6	31.7	4689	6	BD217187
8	158.6	31.7	37948	1	AF079138
9	158.6	31.7	37948	6	BD217173
10	158.6	31.7	38506	6	BD232534
11	158.6	31.7	38506	6	AR271680
12	158.6	31.7	38506	6	AR277664
13	158.6	31.1	4725	6	AX697993
14	155.8	31.1	60196	6	AX697977
15	154.4	30.8	15970	1	AE007061
16	154.4	30.8	348186	1	EX842578
17	153.6	30.7	104326	1	AB070940
18	153.6	30.7	281450	1	AP005032
19	152.6	30.5	65140	6	AX211705
20	152.6	30.5	123580	1	AF263912
21	152.6	30.5	125401	6	AX211739
22	148.8	29.7	11238	6	AX697991
23	147.2	29.4	104326	1	AB070940
24	147.2	29.4	306050	1	EX248341
25	147.2	29.4	320150	1	AP005033
26	144.8	28.9	113193	1	AF357202
27	144.8	28.9	113193	6	AX703543
28	143.2	28.6	107379	1	SHGCCPIR
29	143	28.5	84985	1	SNA278573
30	142.2	28.4	2706	1	AF420594
31	142	28.3	27541	6	AX211706
32	141.2	28.2	2700	6	AX211733
33	141	28.1	78210	1	AB070949
34	141	28.1	300425	1	AP005022
35	140.8	28.1	281450	1	AP005032
36	140.4	28.0	132544	1	AF521085
37	140.2	28.0	11220	6	BD217186
38	139.6	27.9	30690	6	BD097649
39	139.6	27.9	64957	1	AB032367
40	139.6	27.9	302675	1	AP005024
41	139.6	27.9	138203	1	AX310323
42	139	27.4	35445	1	MLCB2052
43	137.2	27.4	39314	1	SGR300302
44	137.2	27.4	318200	1	MLEPRTN6
45	137.2	27.4	318200	1	MLEPRTN6

ALIGNMENTS

RESULT 1  
AMM223012  
LOCUS  
DEFINITION  
53784 bp DNA linear BCT 09-FEB-1998  
Amycolatopsis mediterranei genes encoding rifamycin polyketide  
synthases, ORFs 1 to 5.  
ACCESSION  
AJ223012  
VERSION  
AJ223012.1 GI:2764760  
ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.  
KEYWORDS  
Amycolatopsis mediterranei  
SOURCE  
Amycolatopsis mediterranei  
ORGANISM  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.  
REFERENCE  
1  
Schupp, T., Toupet, C., Engel, N. and Goff, S.  
AUTHORS  
Cloning and sequence analysis of the putative rifamycin polyketide  
TITLE



synthase gene cluster from Amycolatopsis mediterranei

Unpublished  
2 (bases 1 to 53784)  
Schupp T.  
Direct Submission  
Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma  
AG, CH 4002 Basel, Postfach, Novartis K-681.344, 4002 Basel,  
SWITZERLAND

FEATURES  
source  
1. 53784  
/organism="Amycolatopsis mediterranei"  
/mol\_type="genomic DNA"  
/strain="IBG A3136"  
/db\_xref="taxon:33910"  
1336..15543  
/function="CoA ligase, acyltransferase, ketoreductase"  
b-ketoacyl-ACP-synthase, acyltransferase, ketoreductase"  
/note="ORF1  
PKS modules 1-3"  
/codon\_start=1  
/transl\_table=11  
/product="rifamycin polyketide synthase, type 1"  
/protein\_id="CAA11035.1"  
/db\_xref="GI:2764761"  
/db\_xref="GOA:054666"  
/db\_xref="SPTREMBL:O54666"  
/translation="MRTDLIKPLHVALLENATRPAGKPAFADHRTVYGGLEARTRR  
LAGHAGLVGRHGRVRAICLGRVSTVESFPAIRAGAGVPLNPGSATAEHLPLTD  
SGATVYDQAARLARBLAHVELLVTDGDDPEGAHVSVDLASEPAPPAADOLELDE  
PAMFVYSGTTPKPKGVVSTQRCNLSVASCYVPPGLSDQDRVLPPLFHSLSHTA  
CVLSATVAGSVRIADGSSADVMRLLEASSTFLAGVFTTYTHHLVRAAQRGSAPS  
LRIAGLAVGAGVIRGDSFEFTFGVLDAYGSTTCGATITMPPDGRVSGSCLAV  
PDGVRVDPDTGIDPAGGEVGVNPMGLYHNSPEATAAMRDGMFRTGDLAR  
RDDAGYTCIGRIKELIIRGCANTHPGEVEALRTVDGADAAGVGHDTLGEVPA  
YVIFPGPGDPAALIEKREOLSAVKPDRILEVAHPTASGIRKGLITDDEPAQLR  
YAATHEEQSRHABESVAARLARLSGLDERAQCLELLEIVRTQAAVGLGQVPGRA  
FRDGLGTSALVELNRNTEHTGLWLPASAVDFDHTPAALAAARVRAELLGITQVAEP  
VVAADPGPIAIVGACRLPGVSPEDLMLVAERVDVASEFFPDGRGNDLSDLDIPD  
RERAGTSYVGGGFLHDAGEFDAGFGISPREAVAMPQORLLLETLSWEALNEAGVDP  
TALGTDVFGSLMGQGVGAVAPLEGGFTVGVASVAGSVYVILGEPATV  
DTACSSILVAMELAAQRQCSMALAGGVVMTATPGSFVFSRORALADPGCKAF  
AAADGTGWSGEGVVVLERLSVAREGRHILAVLRSVNDGASNGLTAPNGLSQ  
VIRRALAAAGLAPSDVVAEHGTTGLGDDPIEAQALLATYQERKQPLWGLSKSN  
IGHQAAGVAGVIRVQVQALRHETLPTLHVDKPTLEVDMSGALELLTQAKWPRNG  
RPRAGVSFGVGTNAHLILEEAPAEFPAPELVPVLSARSTESLSQAKELA  
SLLEGVSLTEVAGALVSRRAVDERAVVAGSREAVTGLRALTNGSTGTPGVVW  
FPGQGTQAGMRELLASPVFAERIAECACALAPWIDSLVLDREGDGRVDVLQ  
PACFVWVGLAAWESVGRVPAVVGHSQGEIAAACVSGALSLEDAKVALRSQAI  
AELSGRGGSVALGEDDVVSRVLDGVEVAAVNGSVVIAAGDALDALTLELSEG  
IRVRVAVDYASHTRVEDIRDTLAFTLAGISAQAPAVFFYSTVSEVDRDAGLDGG  
YVNRNVRVGRGAATATLEQHTVFVEVSAHPVTVPQSLTGDATGLRREDGG  
RLLASMLFVRGIDVDMVMPAAGVVDLPTAFEHRYHWEPAEPASAGDPLIGT  
VYSTGSDRLTAVAGSRAQPAWVDGLVFNAAVLEAAILRGDLACTPVVGLVVDAP  
VILPRGSREVQLIAGEQORRPTIEVSRADDEPETHAGTLAPAAAVPEPAAA  
GDAITVLTGADADRIHLPALLDRAVIVVGGDILLPSVWTGVSLSAGATVTP  
TATGLRTDPAGQVLTIVSVRGTPFAEQGTDLFRVDWPEIPLTETADFLPYE  
ATSATLGAQWALADPAETVLAVTGCTEPEGAAILWGLVRSAGSEHPGLVLDL  
DIPALPVAVASGEQVRVNGVAVPRLTRVTPROQARPLDPPEGTVLTGCTGTGA  
LTARHLVTHAGVRHLVAVSRGEAPELOELTALGASVAAACADRAQLEAVLRAI  
PAEHPITAVHTAGVLDGVTLETLTDRLATRPRKVDARLDELTRADLAAPVLI  
SSAAGVLPNGQAGYAAANALDALARQNSLDLPVASTANGVWATVSGTEHLGDAD  
LRNQRIGMSGLPADMSGLLDAIATGGTLVAAKTFDAALRAKAGGVPFLLRGL  
APLPRAAAKTASLTIERLAGLEAQEAALLDILVRHAAEVLGHSAGSVHSRTFKD  
AGFDSLTAVELNRLEAATGLTSPAMI PDYKPKPALADHLRAKLFGSANRPAEIGT  
AAABEPIAVAMACRPGGVHEDLWRLVADGADAVTEPPADRGMDTDRLYHEDDPH  
EGTYVNHGAFLLDADAGDAAPFGISSENEALMDPQORLLLETSMELFEAAIDPTTL  
AGQDIGNVAGVNDHSMRHAAGVEGFLTGTGSSASVLSGRVAYHFGVGPATVDT  
ACSSILVAMVQALQRCSCMALAGGVMMVTETVFEFSRQGLAPGRCATFAD  
GADGTGSEGVGLTLELRLSEARQGRHQLVAVRGSVNDGASNGLTAPNGPSQQRV  
IRKALAGISTSDVDAEAGTGTGLDPIEAALATYQNRTEPLWGLSVKSNIG  
HTQAAAGVAGVIRKVMARHGVLPRTLHVDPRPSYVDKASAGAVELLTEARDWVSNHP  
RRAGVSSFGIGGTNAHVLEVAAPITTPQEPAPARFLVPLVLSARTAAGLRQAGRIA  
AFLGDRTDVVRPDAAYALATTRAQLDHRAVVLASDRALQCADLAAFGSGVVTGTVDG

CDS  
15550..30759  
/function="b-ketoacyl-ACP-synthase, acyltransferase,  
dehydratase, ketoreductase, acylcarrier protein"  
/note="ORF2  
PKS modules 4-6"  
/codon\_start=1  
/transl\_table=11  
/product="rifamycin polyketide synthase, type 1"  
/protein\_id="CAA11036.1"  
/db\_xref="GI:2764762"  
/db\_xref="GOA:052789"  
/db\_xref="SPTREMBL:O52789"  
/translation="MANQSWRKNNSAPNEQIVDALRASIKENVRLQENSLAALAAAAE  
PVAVSMACRYAGGIRGDFRVRVSEGADVTGFPEDRGWDVGEGLHYHPDNPCTTY  
VREGAFLODAAQFDAGFGISPREALMDPQORLLLEVSWETLERAGIDPHSVRSDI  
GVYAGVHQDAPDLSGFEGMSLERALGTAGVAGSVAYTLGLEPAVTVDTWCSS  
SLVATHLAQAALRCECSMALAGSTVMTATPGGFVGFARALAFDFRCKSVAAAADG  
SWAEGVLLRLERLSVAREGRHVLAVRGSVNDGASNGLTAPNGPSQQRVIRKA  
LASAGVDVDTVBSHGTVGLGDDPIEAQALLATYQERKQPLWGLSVKSNIGT  
QAASGAGVITKMQVSRHGOQLPATVPTQVDMWAGAIIELLAAGSEWPNRHP  
GGISFSGAGSTNAHMLILEEAPAEFPAPELVPVLSARSTESLSQAKELA  
GVSLADVAGTLVSRKMLSERAVVAGSREAVTGLRALSAGSGLLSGRGSGVP  
GKVWVFPQGTQWAGMRELLDSQSEVFAARIAECETALGRWVNDLSLDEADLL  
DRVDVQPASFAVWVGLAWLASLVEPEAVVVGHSQGEIAAACVSGALSLEDAKVAL  
LRSQAISLASGRVGLVAVSHTRHVEAARDALAEMLGGIRAQAPVFPVSTVSGW  
EALDADLQGVYRNRVFRFGVPAVELIEQGRHVFVEVSAHPVLVPLINELVDDE  
AVTGTTLREDGGYVNRNVRVGRGAATATLEQHTVFVEVSAHPVTVPQSLTGDATGLRREDGG  
MGSATDVSILGLADHPLLGAVLPQSDGLVFTSRLSKSHPLWLAGHAGVVL  
AVTGTTLREDGGYVNRNVRVFRFGVPAVELIEQGRHVFVEVSAHPVLVPLINELVDDE  
SREDDGDEWTRHATGLIGASTSREPRFPAAPAGAPIDVENFTDLTERGYVS  
PGTVVDLALRAGDELFGVLELIEALPLVGERGVVAVVSGNPTGSRVAVF  
GAFQGNRAVVRRGDEVEFAEVLDPDDRDAGKPLHPALDAAHLNANFANPDDESV  
LIPFANGLVLAHVSALRVVAPGPGDEHALTEAFGEAPLEFLTRVLEAVSWLAD  
EFAAGERSLIFQVDMIEVPTATTAETHEALTEAFGEAPLEFLTRVLEAVSWLAD  
AADAERLVVTRGAVPEVDPAGAVKGLVRAQAENPGRIILVDTGDDGLVGLAVARH  
GEPLQAVRAGVSVPRALATGEVPEAPVPEGTVLTGTTGGTGGGLVAKHLVARH  
GVRRLVLRGRGVAAEDLVTELEQAGATVSVACVSDRDQVAAALAEHRPTGIVHLA  
GLLDDGVTGALNRERLAGVFAKVDVAQVHLDELTRDLGLDAFVVFSSAALMGSAGG

CDS

NYAANAFLDGLMAGRAAGLPVGLSLAWGLMEADGLTANLSATDQARMSRGVLPMT  
PAEALIDFDIGLAEQALLVPLKLDLRLTGOATAGGEVPHLLRLGLVRASRRVTPTA  
ASGGGLVHKLAGRPAEEOEAVLLGLVQAEAAAVLGFNAPELAQTRGFSLDGLPSLT  
AVELNRLSALTAAGKPLATLVPDYPTPVALARHLEELGETVAGAPATPVTIVADAGE  
PTAIYGMACRLPGVWMDPDLRMWAEGRDMSPPFGDGLDGLDPSDPERPSTAY  
TROGGLHEAALFDPGFFGISPREALMDPQORLLEASWEALERAGIDPTKARGDAV  
GVFSGLTHDXLESLSNPALEGFVTATATASVAGSRVSYTFGFGPAVTVDTACSS  
SLVAHLAAQALROECTMALAGVAVMSPIGVLMGSRORGMWADGVKAFADGADG  
TVLSBGGLVILVERLSVAKERHRVLAVLGRSAVQDGNASGLTAPNGPSQOQVIRSA  
TAGAGLQPSVDVSEAHGTGTALGEPTEAQAALLATYKGRPTPLWGLSLKSNIGHTQA  
LAGAAVAKIMVALLQDPTLPHVQEPKQVNDVAGAVELLTSGREWARNGHPRRAG  
VSGFGTGNHALLILEEADPTADVDPAVVPVVISARSTGLGQAQRLAAFLDG  
DVELTRVAGALLSTRATLTVRAVWVAGSAREAGLTLARGESASGLVTGAPMGPK  
TVWVPGQGTQWAGRELESPVFAERIECAALOPWIDLSLIDVLRGEGELDRV  
QIAAALSGRGWASIQLSHDEVAALFLWAGRVETIAAVNGPASPVSIVAGIEWITDAGVVD  
EVLGRRVADVASHTRHVEDIQTLLAETLAGIDIAQAPVVPFSTVAGIEWITDAGVVD  
GGVWVRLNQVGFVAGVAELIEQHGCVFVEVSAHPVLQPISELDVAVTVTLRRDD  
GGVRLLTSMALFVRGVPVDMATMAPPAVELPTVAFDQHFWLSPPPVADAPALGL  
AGADHLLGANLPLPOSGLVFTSLRSLVTHPLADGVPAALVELAVLAGDEAGCPV  
LADLVKELVLVLPESGGLRQVIVSGERTVYVYVQLEGAEDWIRNATGHLSATAPAH  
AFDFTAWPAGAQVQDGLMRRGDEIFAEVALPEELDAGAFGHFPLLDRAVQVPLADD  
EQPAEWSLVLAAGASALRVLPGCALQADDETGLVLTADSVAGRELSAGKTRAG  
SLVRVDTVSTIADSAVPANIIVEVAFGEPELTLGRVLEAVQTVLTVTRADDARLVVV  
TPGAVREVTDPAAGAAVGLVRAQAEENPGRIFLIDTDGHEIPALTGDEPETAVRGGKFF  
VPRITRABPSGAVERPDGTVLISGAGLGLVARRIVERHGVKLVASRRORDRLA  
VADLVADLAADSVVACVSDRAQVALLDEHPPTAVHTVAGVIDAGVLETLDRDLA  
TVFAPKVDVVRHDELTDRDLDAFVYSSVAVFVAGSGSTAAANAFLDGLMANRR  
AAGLGLSLANGLDQSTGMACTDEATRAMSRRLGQIMTQAEQMDFLDALSSAE  
SLVLPKALDLVRAGAAAGVPHMLRGLVRAGRAQAAASTVDRGLAGRLAGLAPA  
DQTLTLLDIAVRAOVAVLGHADASVRVDTAFKADGDSLTAVELNRMTATGLKLP  
ATLVFPYPPQALABHRLDELGAOCTPVTTAAAKDLDEPIALVGMACRLPGVAGP  
EDLWRLVAGRDVSPPTDRGNDTSLYDPPDPARPKTYTRHGGFLHEAGLDAGFP  
GISPREAVMDPQORLLEASWEAMEDAGVDPLSLKNDVGVTFGFGQCVVAPGSV  
VTEPEFAGTGGSSVAGSRVYVFGFEPVAVTDSACSSSLVAMHLEAQSIRQREB  
SMALAGATVWMPAGVFEFRSQRLAVDRCKFAFAAADGTGMAEGVAVVILLERUSV  
ARRGHRILAVLPGASVNOVAGNGLTAPNGPSQQRVIRRLVAGSLAPSDVDVEAH  
GTCTTGLDPLEAQAALLATYKDRSLPLSLKSNIGHAQAAGVAVGIVKMQVLRHE  
VLPPTLHVTRTEVDWSAGAVELLTEAREPNRPNRRAGVSAFVSGVGNHILIRE  
APAEFPVTPTEVPLVPPVVSARSASLAGAQRLAGFVAGDASLAGVARALVITNRAAL  
TERAVMVVVSREAVNTNLEALARGEDPAAVVTGRAGSPGKLVVWVFPQGSQVIGMGRH

Query Match 100.0%; Score 501; DB 1; Length 53784;  
Best Local Similarity 100.0%; Pred. No. 1.4e-43;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGGTCCCGACGCGAGCAACCGCATCTCTGCTGAAGCTGGTCCGCGGCCACGCTTCG 60  
Db 15043 GCGGTCCCGACGCGAGCAACCGCATCTCTGCTGAAGCTGGTCCGCGGCCACGCTTCG 15102

Qy 61 ACGGTGCTCGGCGCGCGAGGCAATCGCGCCGCGCGAGGCAATCGCGCCGCGAGGCTC 120  
Db 15103 ACGGTGCTCGGCGCGCGAGGCAATCGCGCCGCGCGAGGCAATCGCGCCGCGAGGCTC 15162

Qy 121 GGCTTCGACTCGTGGCGCGCGTCAACTCTCGCAACAGCTTCGACGCGGCGCGAGGCTG 180  
Db 15163 GGCTTCGACTCGTGGCGCGCGTCAACTCTCGCAACAGCTTCGACGCGGCGCGAGGCTG 15222

Qy 181 CGGTGCTCGGCGCGAGGCTGATCTTCGACTACCCACCCCGGAGGCGCTGGTGGCTTACCTG 240  
Db 15223 CGGTGCTCGGCGCGAGGCTGATCTTCGACTACCCACCCCGGAGGCGCTGGTGGCTTACCTG 15282

Qy 241 CGGTGCGAATCTCTCGGCGCGCGAGGCAATCGCGCCGCGCGAGGCAATCGCGCCGCGAGGCTC 300  
Db 15283 CGGTGCGAATCTCTCGGCGCGCGAGGCAATCGCGCCGCGCGAGGCAATCGCGCCGCGAGGCTC 15342

Qy 301 CGAGTCTCTCGGCGCGCGTTCGGTTCGGCTTCAGAGGCGGCGTGTGCTGACACGCTG 360  
Db 15343 CGAGTCTCTCGGCGCGCGTTCGGTTCGGCTTCAGAGGCGGCGTGTGCTGACACGCTG 15402

Qy 361 CTGCGCTCTCGGCGCGAGGCTGATCGACGCACTGAGCATCTCCGCTTCCTGCAACGCA 480  
Db 15403 CTGCGCTCTCGGCGCGAGGCTGATCGACGCACTGAGCATCTCCGCTTCCTGCAACGCA 15522

```

RESULT 3
AF040570          90445 bp      DNA      linear      BCT 16-APR-2001
LOCUS             Amycolatopsis mediterranei rifamycin biosynthetic gene cluster.
DEFINITION
ACCESSION         AF040570 AF040571
VERSION           AF040570.2 GI:123331604
KEYWORDS
SOURCE
ORGANISM          Amycolatopsis mediterranei
                  Amycolatopsis mediterranei
                  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
                  Pseudonocardiaceae; Pseudonocardiaceae; Amycolatopsis.
REFERENCE
1 (bases 1 to 90445)
August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R., Yu,T.W.,
Taylor,M., Hoffmann,D., Kim,C.G., Zhang,X., Hutchinson,C.R. and
Floss,H.G.
TITLE             Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
                  the molecular analysis of the rif biosynthetic gene cluster of
                  Amycolatopsis mediterranei S699
JOURNAL           Chem. Biol. 5 (2), 69-79 (1998)
MEDLINE           98174059
PUBMED            9512878
REFERENCE
2 (bases 1 to 90445)
Kim,C.G., Yu,T.W., Fryhle,C.B., Handa,S. and Floss,H.G.
3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
formation of the precursor of mC7N units in rifamycin and related
antibiotics
JOURNAL           J. Biol. Chem. 273 (11), 6030-6040 (1998)
MEDLINE           98165773
PUBMED            9497318
REFERENCE
3 (bases 1 to 90445)
Yu,T.W., Muller,R., Muller,M., Zhang,X., Draeger,G., Kim,C.G.,
Leistner,E., and Floss,H.G.
Mutational analysis and reconstituted expression of the
biosynthetic genes involved in the formation of
3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin
biosynthesis in amycolatopsis mediterranei S699
JOURNAL           J. Biol. Chem. 276 (16), 12546-12555 (2001)
MEDLINE           21201076
PUBMED            11278540
REFERENCE
4 (bases 1 to 90445)
August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R.,
Hutchinson,C.R. and Floss,H.G.
Direct Submission
TITLE             Submitted (30-DEC-1997) Dept. of Chemistry, University of
                  Washington, Box 351700, Seattle, WA 98195-1700, USA
JOURNAL
REFERENCE
5 (bases 1 to 90445)
August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R.,
Hutchinson,C.R. and Floss,H.G.
Direct Submission
TITLE             Submitted (19-JAN-2001) Dept. of Chemistry, University of
                  Washington, Box 351700, Seattle, WA 98195-1700, USA
JOURNAL
REMARK            Sequence update by submitter
COMMENT           On or before Jan 22, 2001 this sequence version replaced
                  gi:2792340, gi:2832263.
FEATURES
source            location/Qualifiers
1. 90445
/mol_type="Amycolatopsis mediterranei"
/mol_type="genomic DNA"
/strain="S699"
/db_xref="taxon:33910"
423..1451
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC01706.1"
/db_xref="GI:2792310"
/translation="MAPFGGSGRYKTKLQLEVALIGALILHLEWLGAAV
RYDDGSRLEAEFGAKAGSLEALDGDADVDICTPTASHHEIALTAAGGVV
CEKPLAASAEAEIVTAERAGVRLYAAHDVRAAPVRLHVELVAGSLGELGRF
SFSYHPFPWTHASARSGGILITDQLLGHADLHWFGDVVRVHACYGDLATPAPE
AVATCTAIVTHASGAISSVSRKWTATPPPVVAFVHVSSTGSGSYDSSEWQEV
GGANGFAGGGSVFTEPREFATAPAGGPEPRIGAKDALAAVRIHAAAESAWTGRA
ELPVRGAA"
CDS
1448..2215
/transl_table=1
/product="unknown"
/protein_id="AAC01707.1"
/db_xref="GI:2792311"
/translation="MKVAILSSSTPQVAGALRGLEPVEVVAASWDAPFVPOAAEAG
ARVLCYPPPAKATDLKAMIDADAGRLTFASPACHGAFVVRKIGADIGELITVYL
GSVATSDVGLGAAAYLLDADALGAGEPAQVYQNTNIVLSRIGESAAVILRYR
SGVASPDCRHHGATGLPAVTFIGDQSVQYDAGPQLGGERPELGEDLEALMLKD
FLGADGPGPGQALRTFRILQIAYESAHTGQFVDL"
2353..2619
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAC01708.1"
/db_xref="GI:2792312"
/translation="MQDRSVLGDPPKQVEAAMSVTQKIALSGGPAELAPAHRSVDSG
ALENTKIRHGAGYEHFVHGGEFQAVDGCADIARVWSHRTKIAE"
2652..3842
/notes="ORF0; similar to Streptomyces Sp. ChoP"
/codon_start=1
/transl_table=1
/product="cytochrome P450 monooxygenase"
/protein_id="AAC01709.1"
/db_xref="GI:2792313"
/translation="MTDAISFEVPWDRDTPKDPAPVDFSLREERFLAKRVDPGHVGM
TVSSYELREVLSDLRFSGHSCFVHPVTHQQVPTPLIPMGTHMDPPHTRYRK
LITGEFTVRRASRLIPRAEVAAGIOIEMVRKAGADVMDFAKPIVLMLGSLVGLP
YEERDRYVPAVTLHDHADPAEAAAAYVAGKFFDEVTERROQDDLLISLVTEED
LTQOEELNIVTLILLFAGYETTEGALATGVFALLHTDDQLAARAEPEKLDAALELLR
VLTQVYHTYRTALEDEVKLEGLIKKGDVTVLSLPAANDRPKFGCPAELDTERDTS
HYVAFGFIHQCLQNLARIELRAGETALLRAFPFELRLAVPADEVLPRLKGSVFSVKKL
PVSW"
3883..18090
/genes="rifa"
3883..18090
/genes="rifa"
/notes="Rifa"
/codon_start=1
/transl_table=1
/product="polyketide synthase"
/protein_id="AAC01710.1"
/db_xref="GI:2792314"
/translation="MRTDLIKPLHVALLENATFACKGAFADHRTVTYTGDLERTRR
LAGHLAGVYRGGDRVAICLGNRVSTVESVFAIRAGAVGVLNPGSATALEHPLTD
SGATVVTDAQAARLRLAPHVELLVTDGDDPEGAHSYDELAISEPAADLELDE
PAMFTSGTTPKGVSTQNCIASVASCYVFPCLSDQDRVLPFLPFLSHSLIA
CVLSATVGSARIADSGSADDVRLIEASSTFLAGVPTTTHLVAARQRFSPAPS
LRITLGGAVLGGALRSEFEETFGVLIDAYGSTETCGALTMPDGGARVSGGLAV
PGDVRVDPDGTGLDVPAGEEVWVSGPNVLMGYHNSPEATAAMRDGMFRITGDLAR
RDDAGYFTICGRKELIIRGGANIHPGEVAVLTVGDVADAAGVGVPHHTLGEVPA
YVTPGPTDPAALIEKREOLSAKYVPDRILEVAHIPRTASGKIRRLGLTDEPAQLR
YATHEEGSRHADESVALRALRALSGLDERAOCELLEDLVRTQAADVLQGPVDGRA
FRDLGFTSLAIVELNRNLTEHTGMLPASVDFHPTPALAAARVRELIGITQAVAP
VVAADPEPTAIYVGMACRLPGVSPEDMLRLVAERVDVASEFFPGRCNMLDNLDPD
RERAGTSYQGGFLADGEFDFAGFTGISPEAVAMDPOQRLLELTSWEALNGLGAV
IALKGTDTGVFSGMLGGQSGAVAPLEGFTTVTVASSVSGSVSYVLEGGPVTV
TAACSSLVAMHLAAQALROGCSMALAGGVTVMATPGSFVFSRQALAPDGRCKAF
DAAADGTGSEGCVGVVRLSVAREGHEILAVLGRSAVNDQASNGLTAPNGLSQ
IRVIRALAAGLAPSDVDVVERHGTTLGLDPTLEVDMSAGIELLTERAWPRNG
IGHAQAAGVAGVIMVQALRHETLPTPLHVDKPTLEVDMSAGIELLTERAWPRNG
RPRRAGVSFSGVTNAHLILEAPAEVPAPELFPVPLVVSARSTESUSGQERLWA
SLLEGVSLTEVAGALVSRRAVLDERAVVAGSREAAVTGLRINTAGTSGTTPGVVW
PGQGTQWAGMGRLELAEVFAERIAECALAPWDWSLVDVLRGEGDLGRVDVLIQ
PACFANVGLAAWESVGPDPVAVGHSQGLIAACVSGALSLEDAKVVALARSOALA
AEUSGNGMSAVALGEDDDVSRVLVDGVEAAVNGPSSVVIAGDAHALDATLEILSGG
IYRVRVADVASHTRHVEDIRDTLAETLAGISAQAPVPTFTSVTSEWRDAGVLDGG
YVYRNLNQRVFGAAATALLDQHTVFVEVSAHPVTQPSLTDGALITLRREDGGL
RLLASMGELFVRGIDVDWTWPAAGWVOLTPTVAFERHYWLEPAPESAGVLLGT
VYVTFSGDRDLTAQWQSRAPQAMVDGLVFNPAALVEAAIRGLDLAGTPVWVGLVDFP
VLPFRSGREVQLIIVGEGQRRRPIEVFSEADPEWTRHAGTGLAATAAAVPPPA
GDATDVTVAGLRDADRYGIHPALLDAAVTVVGGDILLPSVWTGVSLLASGATAVTPT

```

TATGRLRLTDPAGQVLTIVSVRGTPVVAECQGTDTALFRVDMPEIIPLETAETADFLPYE		
ATSAEATLSALQALADPAETRLAVVVGDCTEPGRALWGLVRSQSEHPGRILVLADL		
DDPVLPAVAGSEQVVRNGVASVPRITVPRQDARPLDPREGTVLITGTGTGGA		
LTARHLVTAHGVHVLVVRGEAELELTAALRGAVALAACDVAADRAEVLARLAI		
PAEHLVTAHGVHVLVVRGEAELELTAALRGAVALAACDVAADRAEVLARLAI		
SSAAGVLPNGQAGYAANAELDALAORNSLDLPVSLWANGYATWATVSMTEHLGDGL		
LRNORI GMSGLPADEGMALLDAAIATGCTLVAEKEDVAALATKAGGVPVPLRLGL		
ALPPRAAKTASLTSLAGLAETQAALLDLVRRHAENVLCHSGASVSHGSRFPK		
AGFSLTAVELNRAAATGLTLSPAMIFDYKPPALADHLRAKULFGSAAANPABIGT		
AAAEPIAIWAMACFPQGVHSPEDLWRLVADGADAVTEFPADRWGDTLRLYHEDPH		
EGOTLVGFAGNDDHAGDAFAFGISPEALAMDPOQLLLETSELWELFERAAIDPTTL		
AGQDILGVFAGNDDHAGDAFAFGISPEALAMDPOQLLLETSELWELFERAAIDPTTL		
ACSSILVAMHVALQALORCECSMALAGVWMTVETVFRSQRGLAPDGRCKAFAD		
GADGTWSEGVGLLVERLSEARQHQVLAIVRGSAVNSDGSNGLTAPNGPSQORV		
IKDLAAGLSLSDVDAVHAGTGTITLGDPIERAEALLATYQNRETPLWLSVKNLGL		
TRTAAAGVAGVIMWAMHGVLPRTLVHDPSSVYVDSAGAVELLTEARDVNSNPH		
HAGVSSFGITGNVHVLVEEVAAPITTPQBPFAELVFLVLSARAAAGLQAGRLA		
AFLGDRDVRVPDAAVALATRAQLDRAVLAASRAQLCADLAAPGSGVWTGTPVDG		
KLAVLPTGQSWAGNGRELARTFPVFRDAFPAACVADTHLRERPLRWFDDSLAL		
DQMTYQALFAVETALFRLFSWGVRLGSLAGSILGELAAHVSGLDLADAGELVA		
ARGRLMQALPAGAMVAVQATEDEVAPLDDGTVCVAANGPDSVLSGTEAAVLAIVAD		
ELAGRGKTRRLASVHAFSPLEBPMDDFRAVERLTFRAGSLPVSITLGLAALD		
SPDYVWGVNRAVFDADATLGAQASTFLELPGGALAAMALGTGLGPEQSCVATL		
RNGAEVDPVLTALAEHVRGVGDMTTLVDEPATVAVTLPYAFQHORFWVDVDET		
KLNRLDFTGVLPSTLIFDYPNPAVLADHLRAELLGERPAPAPVTRDVSDEPIAT		
VGNSTLPGQASPESELKWLVAEGDVSFPVDRGNDLDDGLVHPDPAHAGTSYTRSG		
GFLHDAQPDAGLFGISPEALAMDPOQLLLETSELWELFERAAIDPTTL		
GIVHDDYVRLRVEDVQYTMGTASSVASGRVAVFVGEPVAVTPTACSSSLVA		
MHLAQAALQEGCSMALAGATWASPDALFEEFSGRLSADGRKACRAEAGDGTGWA		
EGVGVVRLSVAERGRHVLAVLRSANODGASNLGTAPNGPSQORVIRGALGSA		
GLAPSDVVEHGTGTALGDLPIVQALLATYQEREQPLWLSLKNLGHGTAAGV		
GVGIMMAMRHGMVPAITLHDEKTSQVMSAGAEVITEARWPRTRGRPRRAGVSF		
GASGTNAHLIEEGPAAEVEDEVSVPLVSVARSAGSLAGQAQALVAIENESLAG		
VAGALVNSQTRILNERAVIAGSDEAQGLQALARGENAFVVTGTAGKPVVWVFP		
GQSSQWNGRDLNDSVPFAARIKECAALEQWTDWLSLDLRLGADLDRDVLVQP		
ASFAMVGLAAVTSGLTVDVLHSGRQIAAACVSGALSLLDAAKVAALRQALAG		
ELAGRGMASVLEBEDAVAPLWPNRVEVAANDVSPSVVLAQDAALEALAG		
DGVRRVAVDYASHTHVBAIARTLAKTLNAGVAVLVAADVAPLFTVSTVLTGWLQAVDA		
GYNRLNQVRQVPSVADVAGLGHVTVFVISAHPVLVQLSEISDVAVVTGSRDD		
GGLRLILASAEVLYVADVATAVPAAGVLDLTPYAFDRHFWLHEARSTAEKRLQVY		
DGEFTWAIQSDVLSLAELVLEPQORGLSTVVPVLAQWRDRRERSTAEKRLQVY		
WPLEHDTGLVSLALDGGPADAAITASTIALVQALGDTTTSAPLWCLTSVAVNIGI		
QDVAVTAQAAGVGLKRAVALERLDWRGLVDLPALADARTQAQLGVWLAGSAGEDQL		
AVRSYVRRRLVRKVPESATSRWEPRTVLTGGAEGIGLGRHASVWLAGSAGEDQL		
TGTDVDELTAELAEFTGTVEFCADTDRDAIQLVADSEVTA VVHAADIATQTSVDVT		
GVADLDFVAAKTVTAVLDQLEPDTPLDAVFFSVIAGIWWGGGQGPAGAANAVIDA		
LVWERRAGLAKTSIANGALDOIIGIMDEAALQALRRRGTIPMAPPLAVTAMVOAVAG		
NEKAVAVMDMWAIFAPTSVRPSPLPADLPEAKAILRAQDDGDDGTASSLADSL		
RAVDAQNRLKLVHGHASTVLSHSGARGIGPQAFQEVGFDSLAANVLRNLSLHAA		
TGLRLPATLIFDYPNPAVLVYLRVLELLREADDGLDGRDRLVRLVAPFAPRKFAAG		
VLDTLLGLADTGTBFTDAETTEAAPAAADDAELIDLALDISGLVQALGQTS		
3883..5775		
/gene="rifa"		
/note="encodes polyketide synthase loading domain"		
Query Match 100.0%; Score 501; DB 1; Length 90445;		
Best Local Similarity 100.0%; Pred. No. 1.le-43;		
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
misc_feature		
Qy	1	CGGTCCTCCGAGCGCGAGACGACGATCCTGTAAGCTGTCGCGGCGACGCTTCG 60
Db	17590	CGGTCCTCCGAGCGCGAGACGACGATCCTGTAAGCTGTCGCGGCGACGCTTCG 17649
Qy	61	ACGGTGTCTCGGCGCAGCGCGCGCGACGATCGCGCGCGCGCGCTTCGAGAGTTC 120
Db	17650	ACGGTGTCTCGGCGCAGCGCGCGCGCGACGATCGCGCGCGCGCGCTTCGAGAGTTC 17709
Qy	121	GGCTTCGACTCGCTCGCGCGCGCTCAACCTCCGCAACAGCTGTCACGCGCGCGCTG 180
Db	17710	GGCTTCGACTCGCTCGCGCGCGCTCAACCTCCGCAACAGCTGTCACGCGCGCGCTG 17769

Qy	181	CGGTCCTCCGCGAGCGCTGATCTTGACTACCCACCCCGAGCGCGCTGTCGGCTACCTG 240
Db	17770	CGGTCCTCCGCGAGCGCTGATCTTGACTACCCACCCCGAGCGCGCTGTCGGCTACCTG 17829
Qy	241	CGCGTCGAATCTCTCGGGAGCGCCGACGACGCGCTGAGCGGGGGAAGACGACCTCCGG 300
Db	17830	CGCGTCGAATCTCTCGGGAGCGCCGACGACGCGCTGAGCGGGGGAAGACGACCTCCGG 17889
Qy	301	CGAGTCCTCGCGCGCGCTGCTTCCGCGCTTCAAGGAGCGCGCGCTGTCGACACGCTG 360
Db	17890	CGAGTCCTCGCGCGCGCTGCTTCCGCGCTTCAAGGAGCGCGCGCTGTCGACACGCTG 17949
Qy	361	CTCGGCTCTCGCGACACACCGGACCGGACCGGACCGGACCGGACCGGAGCGGCC 420
Db	17950	CTCGGCTCTCGCGACACACCGGACCGGACCGGACCGGACCGGACCGGAGCGGCC 18009
Qy	421	CGGCGCGCGACGACCGCAAACTGATCGAGCGCTGACATCTCCGCTCTCGTGCACGCA 480
Db	18010	CGGCGCGCGACGACCGCAAACTGATCGAGCGCTGACATCTCCGCTCTCGTGCACGCA 18069
Qy	481	GCCTCGGCGACGACGCTGA 501
Db	18070	GCCTCGGCGACGACGCTGA 18090
RESULT 4		
AF506522	22131 bp	DNA linear BCT 30-MAY-2003
LOCUS	Streptomyces hygroscopicus AHBA biosynthetic gene cluster, partial sequence.	
DEFINITION	Streptomyces hygroscopicus AF506522 AF506523	
ACCESSION	AF506522 AF506523	
VERSION	AF506522.2 GI:31211262	
KEYWORDS	Streptomyces hygroscopicus	
SOURCE	Streptomyces hygroscopicus	
ORGANISM	Streptomyces hygroscopicus	
REFERENCE	1 (bases 1 to 22131)	
AUTHORS	Gao, Q. and Wang, Y.	
TITLE	Identification and analysis of two separate AHBA biosynthetic gene clusters in Streptomyces hygroscopicus 17997	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 22131)	
AUTHORS	Wang, Y. and Gao, Q.	
TITLE	Cloning of an aminocyclitol biosynthetic genes from Streptomyces hygroscopicus	
JOURNAL	Unpublished	
REFERENCE	3 (bases 1 to 22131)	
AUTHORS	Wang, Y. and Gao, Q.	
TITLE	Direct Submission	
JOURNAL	Submitted (26-APR-2002) Microbial Pathway Engineering, Medicinal Biotechnology, Tiantan, Beijing 100050, China	
REFERENCE	4 (bases 1 to 22131)	
AUTHORS	Wang, Y. and Gao, Q.	
TITLE	Direct Submission	
JOURNAL	Submitted (10-DEC-2002) Microbial Pathway Engineering, Medicinal Biotechnology, Tiantan, Beijing 100050, China	
REMARK	Sequence update by submitter	
COMMENT	On or before May 30, 2003 this sequence version replaced GI:24637555, GI:24637552.	
FEATURES	Location/Qualifiers	
source	1..22131	
	/organism="Streptomyces hygroscopicus"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:1912"	
CDS	<1..2412	
	/codon_start=1	
	/transl_table=11	
	/product="ShnC"	
	/protein_id="AAP45191.1"	
	/db_xref="GI:31211263"	
	/translation="XDEEDNVLRLTVVSTLRDEGGPRLLTSMALFVRGVFVDM	
	TKFLMAGHAGVLDLPTVAFDRRHWLQDAATDASGASNDADAFWSAVEQTADSLA	

GLLAPDSAGLADALRTVVPALADWRGSRSSAERLYAYTWPLDREVSVPAGEW  
LAVPFGPCATETVTSRVAELIALGAGQLDVPETVPSPATFTGTLTARLSDIAEY  
QPAGVLSLADGQDVIDTVARTLALVQALGDAGVNGPLMCLTRGAVNTGICQTAGE  
PGDAIHWLGRAALHPDRMGWIDILPATADAHTAQYLVGALNGTAGDQJAVRPGFL  
YSRLVRKPSQTPADGMRPHGTVLVGGAEALGIHASLWARGARLITVTTAQ  
PADAVTEQGLAAAGVETTVVSCADARETLARLIAATPREOPLTVVHAADAPWTS  
AVADTGLDTEVAGKVDTAVWIDELFTGTDAAPLDAFVVFSSIAGVGGGGVSG  
AAGVILDALVDRRGRGLAATSIAWGLDGLGMDENAAAQLERRGLPMHVOVAT  
AFEQAREKAVTVMNDMEAFIPAFTSARVSLFADLPEAALASRSSQDAENGDI  
TSSILSDRLVPQAEQNRLLRLVGGQAATVILGSSGESIGLFGQFOEVGSDSIGAVN  
LRNSLHATGLPATLVFDYPTDPAVVGFLRELLTETSDDEGREDLRLVLAQVP  
LSRLREAGLLDTLLSLGSDVDSVPAAAAPEPAAPAAEDAAVIDVMDVADLVKRAL  
GSGNP"

2461..8979

CDS

/codon\_start=1  
/transl\_table=11  
/product="ShnB"  
/protein\_id="AAP45192.1"  
/db\_xref="GI:31211264"

/translation="MTAPDEQIVDALRASLKENWRLQENQRLSESSAEPIATVSMAC  
RYAGIRNPELWRVNDGTDTYTSFPENRGWDLGEGITHPODPNPGTYTREGAFIHD  
ANLFDAGLFGISPLEALAMEPQORLLEICWEALERAGIDHSVRGADIGVYAGLHQ  
DYAPDLGLEGLTSLERALGAGSAGTASGRVATTLGLEPAVTVDMTCSSTLVAVHAT  
QALRRGECAMALAGATVMTSPGGFIGFARQALAFDGRCKSYGAAADGSMWAGAV  
VILLESARNGHEVLAVIRGSALNQDASNGLTAPNGPAORRYIRKALENAGLTVA  
DIDVEGHTGTVLDPLEAQLJATYQDRPEGEPLMLGSVKVIGHTSGSGVAGL  
INAVQALRHGVMPATHVDNAENPQVDSAGAVELLTEARWPELORPRRAGVSSFGAS  
GTNAHMLLEQAPPEASPSAPADGVVPLVISAATASLTGQAEIRGSLFSEASGTV  
ALADVAALITGRASLAQRAVVVDSPEEALAGLALGARDGVRGAVAGGVRSGDGG  
KVVLVPPGGSFVWGMGRLELCESEVFAARVECAVALERWVDSLVLDVLRGDCPVEF  
FEREDVROPASFAVWGLAAVWESVGVVADAUVGHSGEVAAACVSGALSLEDAVRVV  
AVRSKISGLVSGRGMASVGISSEEAVALQOMDGRVEIGAVNSPSSVATADTEAL  
DPAIETEDQGVRYRIADLYASHRSHVAGVQELINEAFDIRSQAPVTFSTATGE  
WTRFAGALDGVYRNLRKSQVRFPALADLGHVTFVSSAHPLVPQPISEVAVGA  
EAEAVTGSRLRHGGPRRLFTSMADLFVYRGTHVDSVGLAACGADARVDLPTAPDH  
KNYMELAGTANDVASLGLSGADHPLLAGVAVPEVPTSGVLCSTRLSLRTHPWLADHAY  
GSAVLVPTALVELVRAGEVCGTLEBELVETPLVPAQGMVRVPAVAGVPEENGA  
RSVAVYSARDDGGTGFIDGTHAXGTLTAAVPAKXFDFTVWPVGAERVKFDEAAE  
FYEDLADRGYVYGAFQGLRGVWRGEEVFAVSLPDEORGEASGYGLHPALLDAHL  
SGFQVPAGEGIRVPFAWNLRLHAAGSELVRITVPFGPDVAVLSQALQADGAGGVLS  
LQSLISRPVQDLRMDTDDGRDALFEVDWTELPAPAGAPLPFSWTLVASTEDVTE  
LAGATEVPAVALQALGGDGHADALITVRLEIVQAWPAAADLAESRLVLTQGVAP  
ADGEDVSPDGAAGVWGLVRAQAENPDRIVLDDLAANSGLDLAGMGSVLTVLAAV  
DPQVAVRTTISLVRPLVAGKRALXIPEAKSEWLGVTGVTGTLDSLXLXPEXXXA  
PLAAGQVRVLRGALGNFRDVLIALGMYPGAAXKXGEGAGVLEVEGVTGLVAGDEV  
MGLTGIGSMTVDHRCVLRIPQGWSPQAAATVPVFLTAIYIGLRLDLAGLAGESVL  
THAAGVGMAATOLARHIGAVFTGASPGKVGWLELDFDXAGTASSABTEFEBHL  
TGTSGAGMDVLDCLAGELVDASLRLPRGRFIEMKGTDIRDAGVSAHEPGVAYQA  
FQVFDAGLDRIOQMLLELARLFTGDLTLPFAAMDVRAPDAFRHMSQARHTGMVL  
SVYPELIDEGTVLTGCTGSLGAVARHLVGEHGVPHVLASRRGPDASGARELVABL  
GEAGATVAVACDVTXBAKXALLAAXPAEHPLTGVVHXAGVLDGVIOTLTPERVA  
VFAKVTAVRHLDLHDMLSNFAVSSAAGVFGAGQNGYGAANAFLDGMWARRRA  
EGLPGTAMAGLMEQTGTMTANLSQADQARMGRGILLPIGVAEGMELFDTLRLMPTAL  
AVPIKDLRGMRAADAGGSVLALLRALVQTRQQAQAGSNGSLTARLASLTPOQE  
ELLDDLVLTQAATVIGVTEISNVAHKAENEAGFDSLTSGVSLNRLREATGLKLAATS  
VEDYPTPLALARHLHDLKPAQSPCTHPLLAELSLEARLVNITETPVDDTSIRAQVTTTL  
QVLLTWTSTANGTPTGDEALDFATSDDDLFLDLITFSN"

9111..13136

CDS

/codon\_start=1  
/transl\_table=11  
/product="ShnB"  
/protein\_id="AAP45193.1"  
/db\_xref="GI:31211265"

/translation="MLELEAKSQPVAIVGMACRFPDGVASPODLWRLVSEGREGLSG  
FPFGRGWDVDFADPXRAGKSYTDRGIFHGAGLFDAGFFGISPREALAMPQORL  
LLEVSWEALERAGIDPGLTGSVDVSGVITQYGMRGETPTELGLXRTGTAXXSVF  
SGRVSVVFGPEGSPSVDTACSSLVAMHLAQALROGECSMALAGATVWNPXTFV  
EFSRQXXSDGRCXASAVADGTGWAGGVVLELSEARNHQVLAVIRSSAVN  
QDGSNGLITAPGSPQQRVIRQALASGLSAADVVEGHTGTGVLDGPTPEAQALLAT  
YQGEREPERFLWLGSSNLIGHTQAAGAVAGVITKWVLEALRGVWPATLHADTPSSQVD  
WSAGAVELLTEARDWPTQPRRAVSSFSMGTNAHLILEHSEFPQAEVSAVFPVSDV  
VPIVVSARSARSLVGOQRLASFVGSASDVALPEVAGLVSRRAVLSERAVVAVAGSRD  
QALSGLALARGESHFGVVTGGSLGGVAGSKVLVPPQGSFWVGMGRLELCEASVFA

ARVGECAVELERWDMVSLVDLRGDCSVFFEREDVRQPASFAVMVGLAAVWESGVV  
ADAVGHSGEVAAAACVSGALSLEDADVVRVAVRSQIISGVLSEGRGMAVSGLSBEAAV  
ARLQOQDGRVEIGAVNSPSSVLTSGDTEALGOVLDALAAAGVARRMEVDPAHSROV  
EAEELIREAPADIRAAQAPLPIVFTSYTGWVEAGVLDGQYVWNLRGVFCGPAIA  
DLADGHTVFFESSAPALVWLINETADANTEVILVSGSLRFGGGLHRLFTSWADLF  
VIRVDMVSGVLAAGADARVDLPTIYAFEHVHWPMTSTDAASLGVADGECGCPHLLGA  
VVAVPTGCVGLTWSLSTRTHPLMADVGVVLPVPGTXLXELVVRADGECGCPHLLGA  
LVTBPLVVAQGXRVQXAVGGPEENGXRSVAVSARDDBGXTGDKGWTRHAAATL  
TAAAPADGDFDFTVMPVPGAERVSFDVAFVGYEEMAGRGYVYGPAGQLRGVWRGEEV  
FAEVALPDEQHGEASRGLHPALLDALQSLVLPADAVDMRVVPEAWNGLRLHAAGA  
SELNRVTPGSDPDAVSLQALDAGCPVLSLESLVARAVDVQOLDRMATDGRDALFEV  
DMSLTPASSVESLPPSLASAEVDTLADAAVAVPAVLEAVVGGDGHADALATV  
RVLEVVQAFWAAAGLAESRLVVTTRGAVPVGEGNVADPAGAAVWGLVRAQAENPDR  
VILVLDLAADVDMGSLVLPSCXG"

CDS

13847..14533  
/codon\_start=1  
/transl\_table=11  
/product="ShnB"  
/protein\_id="AAP45194.1"  
/db\_xref="GI:31211266"

/translation="MPLSPAKESKMTAEKLLKSLKLVTAELHDLRQOVARRTDEPV  
AIVGMACRLPGCVSPQDLWRLVSEGREGVSGFPEGWGVGELFDADPGRAGKSYTD  
RGGLHAGLFDAGDFGFLSPREALAMDPOHXLKALCCGCTGFCGCLPRWRCSWTCG  
GCVLQRPAGVFPCCVLCWCIPIVGGRRRPPGATVELPRGSPVSHRRSRHCCSTWSGH  
RWRSSATSR"

CDS

14963..15505  
/codon\_start=1  
/transl\_table=11  
/product="ShnB"  
/protein\_id="AAP45195.1"  
/db\_xref="GI:31211267"

/translation="MPLSPAKESKMTAEKLLKSLKLVTAELHDLRQOVARRTDEPV  
AIVGMACRLPGCVSPQDLWRLVSEGREGVSGFPEGWGVGELFDADPGRAGKSYTD  
RGGLHAGLFDAGDFGFLSPREALAMDPOHXLKALCCGCTGFCGCLPRWRCSWTCG  
GCVLQRPAGVFPCCVLCWCIPIVGGRRRPPGATVELPRGSPVSHRRSRHCCSTWSGH  
RWRSSATSR"

CDS

15358..16812  
/note="polyketide synthase type I; PKs"  
/codon\_start=1  
/transl\_table=11  
/product="ShnA"  
/protein\_id="AAN63812.1"  
/db\_xref="GI:24637553"

/translation="MGVLADVHASGEPQVALRSGAVLVRLARVADTDRSTRRLDP  
DGTALITGCTGALGALIVARHLVVEHKIRSLVLSRRGPDAPGAADLDELTAIGARVR  
IVACIDREAGAGELIASVPRDAPLTAVVHTAGVLDGVTALTPTERLDVATPRKADA  
ALVDELTHLDVAAFLVLFSSAAGTFGNPGQCNLAASNAYLDALARRRTAGLPATSV  
ANGWQDTGISDGLGVADQRRWRWGLAAHSAQEGLELFDALRAYDAVLVAAXNFA  
GURAQSEFVHVILRGVLRARRAAQSSREGGLAGLQAAATPVQRQIILLDLVRR  
EYAAVLYSTPRKVDPRAFQDVFTSVLAVELRNRLAGLIGLRIPASIAFDPTPRR  
MMRHLLAELCPEDGSEPADREDEIRRALATPLSRFRELGLMEQILHLVAHPSGESAA  
APUTAPKQDAGPLTAEMDIDNLVKRAKEKAKP"

CDS

16813..22131  
/codon\_start=1  
/transl\_table=11  
/product="ShnB"  
/protein\_id="AAN63813.1"  
/db\_xref="GI:24637556"

/translation="MDARONETRIAVASSKVEALRASIMENERLESEVQSTRDSLT  
EPIAVGMACRFPFGVSSPEELWELLADRSVAEAPPTNRGMDLENLDPDLRPGTT  
VYREGAFIHDAGFEFDAGFFGISQSETWMDPQORLMLETSWNEAFERAGIDPAMRGKN  
VGVFAGMAAGQYGTAFHSIPDELEGVYMTGLASVLSGRVSYTFYFEGPAPVITDITAC  
SSSLVALHMAAQSLRSGESSLALVGGTNVMTPTTAVLTARAGGLKDGRCRKAFAASA  
DGTNWAEGVGLLLERLSDAVNGREVLVRATVNOQDASNGLAAPNGPSQQRVIR  
QALAGLSPADVDIVAEHGTCTALGDPTEAOALLTTYGNRAPGLPLWGSVNLG  
HAGAAGVAGVILKMWMRHGVLPRTLHVDETPTEVDWSAGAVELLTEAHWPVGRP  
RHRGSGFGASQDNHAILLEQATEPFSGNLDEKARVLDGSDVVPLVVSARGKAGLAGQ  
ARHLSGLTQRQDTDLVIGSLVRSGPLQNRVLAADREALAGLADARAEASAP  
GVVTPGAESTVGRVTLVFPQQTQWAMGAELELEASFVFAARETECAVLDPITGWSL  
LDVVRQVEGASLELDVLPQVSWALMVSLAEACVGPVPAVVGHSIGEIAAACYA  
GALSLPDARLMVHSRIAEARLVGSGMASITADVKVSSILIEEWPGLIEIAAVNGPA  
GALSLDPAARLMVHSRIAEARLVGSGMASITADVKVSSILIEEWPGLIEIAAVNGPA  
SRVPLSYTVEGRHLDTEFEDVEHFRNMAQTVQDFPAIRSLDVGHGVTIEVSAHPVL  
TSSVQVDLADLEAGPAVVTGLTRDDGGPRFRFLASLAHLHTHGRVVRVDMWEALVGRGEGQ



Query Match 55.0%; Score 275.6; DB 1; Length 22131;  
Best Local Similarity 74.1%; Pred. No. 2.6e-20;  
Matches 366; Conservative 0; Mismatches 119; Indels 9; Gaps 1;  
PVDLPTAQRQRYWLETAESRGAPELGLGVANHPPLGAVTEIPGSDGVLFTSLRLSL  
RTHPVLADHAGAVNLPGCAAFVELAVRADEAVGYGLVCLVLERPLVLPESGVQVR  
VWVEPEDESHRTVQVHSRREAGSRGSTRVHSGLVPEDEGQREDFLTQWPPGATA  
VDPDAFAHYDLAEAGYHYGPAPQFQRAAWTRGEEVFAEVSAPESAGKADEYGLHPA  
LLDAAMHTSLFRPLSDSPKALPFPVWRDLRHADGASTLRVHLPLAPDPTIRLHLA  
DTSGTPVAVSDSLVLRVPELLRVGSGAAKQDFRVAWEPI SVRSVDDDEKAVRVT  
AEDVRAAATAAPVLLLDVAGDRTDPDAARDLSGRVLEAVQAWLAPEAFQDFTVLLAL  
THSGAARVDGDPVPLDLAVATAAGLLRAAQSENVGRILLVDTGDEASARLPLVLAAG

4 GTCCCGGACGCGGAGCAGAACCGCATCTGCTGAAGCTGCTCCGCGGCGACGCTTCGACG 63  
1903 GTCCCGGACGCGGAGCAGAACCGCTCTGCTGCTCGGCTGGTCTGTGGGCGACGCGCGAC 1962  
64 GTGCTTCGCGCCACAGCGCGCGGAGGCGATCGCGCGCGCGCGGCGGCTTCGAGGAGGTGGC 123  
1963 GTCTTCGACACAGAGTGCGGAGAGCATCGTTCGCTCCAGTCTTCACGAGGTGGC 2022  
124 TTGCACTCGCTGGCGCGGTCAACTCGCAACAGCTGCGACGCGCGCGCGCGCTGGC 183  
2023 TTGCACTCGCTGGCGCGGTCAACTCGCAACAGCTGCGACGCGCGCGCGCTTCACG 2082  
184 CTGCGCGGACGCTGATCTTCGACTACCCACCGCGGAGCGCTGCTCGGCTACCTCGC 243  
2083 CTGCGCGGACGCTGCTTCGACTACCCACCGCGGAGCGCTGCTCGGCTTCCTCGC 2142  
244 GTCGAATCTCTGCGGAGCGCGACGACGCGCTGCGAGCGCGCGGAGAGACGACCTCGCGCG 303  
2143 TCCGAGCTGCTGACGGAACAGCGACGACCTGAGGCGCGGAGGAGACGACCTCGCGCG 2202  
304 GTCTTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 363  
2203 GTGCTCGCACAGTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2262  
364 GCGCTCGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGG 423  
2263 AGCTTGGGCGACTCGCTGCGACGCGCTCGCTGCGGAGCGCGCGCGCGCGCGCGCGCG 2322  
424 -----GCGCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 474  
2323 GCGGCG 2382  
475 CAACGAGCGCTCGG 488  
2383 AAGCGCGCTCTGG 2396

AF521897 6567 bp DNA linear BCT 10-MAR-2003  
Streptomyces hygroscopicus ansamycin biosynthesis gene cluster,  
partial sequence.  
AF521897  
AF521897.1 GI:28894473  
Streptomyces hygroscopicus  
Streptomyces hygroscopicus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 6567)  
Wang, Y. and Gao, Q.  
Cloning of ansamycin biosynthetic gene cluster, partial sequence  
from Streptomyces hygroscopicus 17997  
Unpublished  
2 (bases 1 to 6567)  
Wang, Y. and Gao, Q.  
Direct Submission  
Submitted (17-JUN-2002) Pathway Engineering Dept., Institute of  
Medicinal Biotechnology, 1# Beijing Tian Tan Xi Li, Beijing,  
Beijing 100050, China

Location/Qualifiers  
1. .6567  
/organism="Streptomyces hygroscopicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1912"  
<1. .2414  
/gene="shnC"  
<1. .2414  
/gene="shnC"  
/codon\_start=3  
/transl\_table=11  
/product="type I PKS"  
/protein\_id="AAO61220.1"  
/db\_xref="GI:28894474"  
/translation="EEDDNVRLRTVVGSTLRRDEGGPRRLTSMALFVRGVFVDMTK  
FLMAGHVDLPTVAFDRHYWLODAATADSGSDNDADAFWSAVEQTDASLGL  
LAPDSGLRALRTVVPALADWGRSRRSSAERLVAWTRPLDRVSRVPAGRWLA  
VLPPCPAETVTSRVLAELIAELGAQGLDVVPFETPSAFTRTGLTARLSDIRAYOP  
AGVLSLLALDGEQVDTIVARTLALVALQADGAVNGPLMCLTRGAVNTGIDTAGEPG  
DAATLWGLRAALAEHPDRWGLIDLPATADAHATTAQILVGLNGTAGDQLIATTAQAP  
RLVVRKASOTPADGWRPHGTIVLTGCAELGTHASLWARSAGRLIVTTTAAQPA  
DAVTELOGLAAAGVETTVVSCADARETLARLIAETPREQTLAVVHAADAPMTSAV  
ADTGHADLTVEFAGKVDTAWMLDELFTGTDAAPLDAFVFSIAGIWMGGGQGVSGAA  
GADJLALVRRRGRGLAATSTAWGALDGLIGMDEAAALRRRGVLPMAQVATF  
EQAAAREKAVTVADMDEAFIPAPTSARVSPLEADLPEAAALSSQSPDAENGDTIS  
SLVSLRDVPOAEQNRLLRLVCGAATVLGHSSGESIGLPQSFQEVGDSIGLVNLR  
NSLHVATGLRLPATLFDYPTPDVAVGFLRSELTTETSDDLEGGEDLRRVLAQPLS  
RLRAGLIDTLTSLGDSVDGSPVEXAPEAPAPAAEDAAARDRDRGRPRKARSQ  
QPQLT"  
2458. .5433  
/gene="shnd"  
2458. .5433  
/gene="shnd"  
/codon\_start=1  
/transl\_table=11  
/product="type I PKS"  
/protein\_id="AAO61221.1"  
/db\_xref="GI:28894475"  
/translation="MTAPDEQIVDALRASLKENMLQENORLSSESSAPPIAIVSMAC  
RYAGTIRNPEDLWRVNDQTDVTSYFPENRGMDLEGIYHPDPDNFGTTYREGAFIHD  
ANLFDAGLFGISLEALAMEPQOROLLEICWEALERAGIDPHSVRGADIGTVAGLVHQ  
DYAPDLGEGLSLALGAGGTASRVATLGLGPAVTDVTCGSTLVAVHAT  
QALRRGECAMALAGATVNSTPGFGIFARQALAFDGRCKSKYGAAGDSWAEAGV  
VILLESIDARNGHRVLAIVRGSALNQDASNLTPAGPAPORVIRKALENAGLTTA  
DLMVEGHTGTGLDPTTEAQTATATYQDRPEGRPLMWSVKSVIGHTQSSGAGL  
INAOALRHGVMPATRVHDAPNPQVDSAGAVELLTAAATASLTQARLGSFLEASGTV  
GTNAHMLEQAEPEAPSPAPALDGVVPLVISAATASLTQARLGSFLEASGTV  
KVLVFPQGGSPWGMGRLEECSEVFAARVGCACALERWYDMSIADVLRGDCPVEF  
FEREDVQFAPFANVGLAAVMSVGVADAVVHSGSGVEAAACVXSALSLEDAVRV  
AVRSKTI SGVLSGRGMAVSLSEEEAVARLQOQMDRVEIGAVNPSFSSVAITADTAL  
DEAIFLTDQGVRRRIAIDYASHRSHVGAQVEILNEAFADIRSQATVPFLSTATG  
WTRAGALDGSYVNLBSOVRFGPAIDLLADGHTVFVESAHPLVQPISEVVAGA  
EKAVVYSLGRHRGCGPRLLTSMADLVFVGTHTVDWSVGLAAGADARRVDLPTAPDH  
KNYWMELAGTANDVASLGLSGADHPLLAGVAVPVPETSGVLTCTSLRSLRTHPLADHAY  
GRCRACPHCFGCGVGAAG"  
5159. .>6567  
/gene="shng"  
5159. .>6567  
/gene="shng"  
/codon\_start=1  
/transl\_table=11  
/product="type I PKS"  
/protein\_id="AAO61222.1"  
/db\_xref="GI:28894476"  
/translation="MPASTFRTPRLITTTGHSWVPPTMSPRSAAGLITRCWVRW  
FRERRAECCVRGCRFHFIRITLWAGVVLVPGTALVELVVRAGDKVCGCTLEEL  
VIETPLVPAQSMRVQFVAGGPEENGARSVAVYSARDDDGRTGIDGWTWRHAGTIT  
AAVADGDFDFTVWPGVGAERVSDAFYEEEMAGRGYVYFAGFGLRWRRGEEVFT  
AFVALPDECHGEASRGLHALLDALQSLGVRPADGVMDRVPFAMWNGRLHHAAGS  
ELURVTPGSPDAVSLQADAGGPPVLSLESLVRAVDVQELDRMDTQDRDALFEVD  
WSLSPASSVESLSPFSLVASEDVTDLADAAVPAVAVLEAVGEGGSHDALALTVR  
VLEWVQAAAGLAESRLVVVTRGAVPVGEGNVADPAGAVWGLVRAAQAENPDR1



ORIGIN	VLLDLAADVDMGVSVPVLA"	
Query Match	53.7%; Score 268.8; DB 1; Length 6567;	
Best Local Similarity	73.1%; Pred. No. 1.9e-19;	
Matches 362; Conservative 0; Mismatches 123; Indels 10; Gaps 1;		
Qy	4 GTCCCCGACGCGGACGAGCAACCGATCTCGTGAAGCTGGTCCGCGGCGACGCTTCGACG 63	CDS
Db	1899 GTCCCCGACGCGGACGAGCAACCGTCTCTCTCCGGCTGGTCTGTGGCGAGCGCGGAC 1958	
Qy	64 GTGCTCGGCCACAGCGGCGCGAGGAGCATCGGCCCGCGCCGCGGCTTCACAGAGTCCGC 123	
Db	1959 GTCTCTCGGACACAGCAGTGGGGAGAGCATCGCTCGCTCCAGTCTCTCCAGGAGTCCGC 2018	
Qy	124 TTGACTCGCTGGCGCGGCTCAACTCCGCAACAGCTCGACGCGGCGGCGGCGGCGGCGG 183	
Db	2019 TTGACTCGCTGGCGCGGCTCAACTCCGCAACAGCTCGACGCGGCGGCGGCGGCGGCGG 2078	
Qy	184 CTGCGCGGACGCTGATCTTCGACTACCCACCCCGGAGGCGCTGCTGGCTTACCTGGCG 243	
Db	2079 CTGCGCGGACGCTGCTTCGACTACCCGACCCCGGAGCGCTGCTGGCTTACCTGGCG 2138	
Qy	244 GTGAACTCTCGGAGGCGGACGACGCGCTGAGCGGCGGAGAGAGACCTCCGCGGCA 303	
Db	2139 TCCGAGCTGCTGACGGAACAGAGCGACGCTGGAAGGCGGCGGAGAGACGACCTCGCG 2198	
Qy	304 GTCTCGCGGCGCTGCTCGCGGCTTCAAGGAGCGGCGGCTGCTGACACGCTGCTC 363	
Db	2199 GTCTCGGACAGTCTCCGCTCTCCGCGCTTCCGAGGCGCGGCTTCTCGACGCTGCTC 2258	
Qy	364 GGCCTCGCGGACACGCGGACCGAAACCGGCGCGGACGCGGCGGAGACACGAGCGGCGCG 423	
Db	2259 AGCTCGGCGGACTCGCTGCGGCGCTCGCTCCCGGAGCGGCGGCGGCGGCGGCGGCG 2318	
Qy	424 CGCGCGGACGAGCGAGA-----ACTGATCGACGCACTGAGCATCTCGGCTCGT 473	
Db	2319 CGCGCGGCGGCGGCGGAGCGGACGCGGCGGATCGATCGATGATGCGGCGGCGGCGGCG 2378	
Qy	474 GCAACGAGCCCTCGG 488	
Db	2379 AAGCGCGCTCGG 2393	
RESULT 6		
AF007101	32870 bp DNA linear BCT 13-JAN-1998	
LOCUS	Streptomyces hygroscopicus putative pteridine-dependent	
DEFINITION	dioxygenase, PKS modules 1,2,3 and 4, and putative regulatory	
	protein genes, complete cds and putative hydroxylase gene, partial	
	cds.	
ACCESSION	AF007101	
VERSION	AF007101.1 GI:2624946	
KEYWORDS	Streptomyces hygroscopicus	
SOURCE	Streptomyces hygroscopicus	
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
	Streptomyces; Streptomycetaceae; Streptomyces.	
REFERENCE	1 (bases 1 to 32870)	
AUTHORS	Ruan,X., Staszi,D., Lax,S.A. and Katz,L.	
TITLE	A second type-I PKS gene cluster isolated from Streptomyces	
	hygroscopicus ATCC 29253, a rapamycin-producing strain	
JOURNAL	Gene 203 (1), 1-9 (1997)	
MEDLINE	98085969	
PUBMED	9426000	
REFERENCE	2 (bases 1 to 32870)	
AUTHORS	Ruan,X. and Lax,S.	
TITLE	Direct Submission	
JOURNAL	Submitted (05-JUN-1997) 47PAP9A, Abbott Laboratories, 100 Abbott	
	Park Road, Abbott Park, IL 60064, USA	
FEATURES	Location/Qualifiers	
source	1..32870	
	/organism="Streptomyces hygroscopicus"	
	1..32870	
	9637..14925	

/mol_type="genomic DNA"	
/strain="ATCC 29253"	
/db_xref="taxon:1912"	
complement(108..1130)	
/note="orf15; similar to RapK"	
/codon_start=1	
/transl_table=1	
/product="putative pteridine-dependent dioxygenase"	
/protein_id="AAC38060.1"	
/db_xref="GI:2624946"	
/translation="MNFSSLVNLGTSYFENGARVVPVGVNRNLGVNVNVAACEYPT	
LDHGYPELEINMVAPTAEFAFVWVTDABSEHGERDGTIYAHDEYFEFCARVPTGR	
YTEATRAAYTMFELBEEFGYSVFRMNFIDINRDNAEGMEYVDFCRGAFEFQ	
CRLEFDOPPAATGSGGGATFYLLACRSGHVHINPRQVPAHYPRYGRPAFRF	
ARATYLSRADNGVGGQVFSGTASTVLGHETAHEGDLVKQCRLENIELVLSGNLA	
AHGISAGHGTALRNKIVYVRSSEDPVAVREICREAFSPDADIVLYLVDVCSLLVE	
IEGVVM"	
1454..9625	
/note="orf1; loading domain"	
/codon_start=1	
/transl_table=1	
/product="PKS module 1"	
/protein_id="AAC38061.1"	
/db_xref="GI:2624948"	
/translation="MDLRRLRPLPELLKHAARRGSQTAFADSRGRVSYAQLRATR	
RLAGHLAQLRLPGDRAIYLNCVETIETSYLAARASAGVPINHSSTAEALYLD	
DSAARVTVTHTRLPQVRRLQAERPHLTVVVTGDESGAVFDALAEFEPQPPARDLL	
GLDDIANMLYTSGTGRPKGVLSQSRCLASVAASAGVLGSLSDTRVWVPLFHSI	
AHICYLSVTAVGATARI TEAFDABELLHLTRAEPTFLAGVPTMYHYLLDAARDGL	
SLPNRLVLSAGASTATLRABFQETFGVPLDLCYSTETCGLITAMNPNTQAEGSC	
GLVPVGLTVIPEFTHESVETBEIGIIVWGPSLMVGYHNPQTEALPRGWRTG	
DVLRADLDGLYHITGRLESELIIRSGENIHPTVEQVLLRVPGVTDAAVGSAHPTLGE	
PLGAPVPAADGPDDEFAACRLAYFKVPELYEIGSPRTGSGKI KHALLAOSP	
ARLAVSSGSPESLFLWETAPVPEAGVDQDATSAHRAGSVTAGLEAPGSVAVDC	
PRGSLDEVLRVTRGLTDRSRVAVLTHGADHPAQAAVWGLVSRMADADHPAPLV	
DADGEATALASVLASGQPRALVGGVILVPLARAGSGQRVSRPDGVTVLVFGAD	
QEQAAELAGHLVAAVYGRHVRLLTAVPDLVARLTELGA SVTQTRAPGRASLAE	
FQTAGPVTAVLQVAGPETLTTLTGVEGLPLQDALDELTAHPALPVLAPID	
NGPAHVQAVANRRHAAGLPALASWGPRLTSRRLAFDAATVDDASVLAIH	
ELASAVDTGAAGELARLPAAERGRALLDLVSEGVITGGRPVHQDSQSPDICTST	
SAVELRGLAAATGLALPATLAFDYPNPDALAAVLGRLFGDTSAAGDTEPTDPAAP	
SGEPIALVGMACRLPGVTSPEELWELLSSGGDGSIGFFTDTRGWDLHLHPDHPHG	
TSYVREGFLHDAGDFDAALFGISPREALAMPQORLLETTLTWELRAGIDPTSLG	
SRTGVFAGVMYHDIYSLRTVPELEGVLSGGSGSVSRVSTFGPEGAVSDTA	
CSSLVALHLAVQALRTGECTLA VAGVTVWASPSVYIEFSRQRLSPDGRCKSFAAD	
ADGAWSEGANLTLVERLSDAQRLGHPVLGVVGTAVNODGASGLTAPNGSPQORVI	
RDAUTNAQTATDIDLTEHGTGTGUGDPIEAQVLA MYGQDRDLFWMLGSLKSNIGH	
TQAAAGVAGVNAKVLNMRHGVMPKSLHIDEPTPEVDWSAGTVLLAEARDWPGADERP	
RRAGVSSFGVGTNAHVILSQAPPAPEPTEAEFESEAGSPGSPVGLGSPGPVP	
WVLSAQNETALKOLARQLVTEGTGHIASHLTQSRSTTLPHRAVLVGSREBFATA	
LEQITRTPDGGKTVLVPGGOSQWGMQORLWETSPLYRDTHEIDTALSEYIDWSV	
TDILTSORGAPTDVVDVIOPLVFTTWTALATLWQHGLTPDAIIGHSGRIIAAYTA	
GALTQDAARITTLRSKALLPLTGHGTWISVLAPPEKIOELADPDHNRNLIWAHNA	
PNITTIYTDQTAITLTLTQOLSHRMRWQLPGVDFAGSHIDQLHQLKILAPITP	
QKPTIPFYSTDNTWITDITDADYWRNLQFVLPVFPAINHLNNTYTIETSPHP	
TLTPSIQETNPNTTIHTLNNQNDHALTALGHATHGHPTIHTLIPITKTPTLP	
TFPQYTRYLNDKPTVTDQATOLGILPTPLVPAATTLTAETGATLITGHINPAH	
PWTTDHAVNGTELLPGTALVDMALHAGDHTDHTLDELITHTPTLTHPTTITQITSP	
PDNNTHTLVTHRNSPDGWAHATGTLTQHQTHTPDPPTQTTWPTPEAPLPLDGTQ	
RLATTIGLYGTFOGLDMLWRDRLFADVLKPNTHDSDAGGIGHPALIDAAHAL	
VASVDGEIRLFPFAWGTTLHATKRLVLEWDGSGASLAFSTFGQVDDVTESL	
ASRVISPEQLRSASGLPYRPEWIAWSAGTGGAVFEYEVPTFGQVDDVTESL	
RSTETLTRIQQHLLAADDTPLVVATHDLLAGAAVWGLIRTAQTEHPRGRIILLIDNH	
PTSEKALPTLIASGEPQARIRDGAILTLPRLTVVATPAPGAGVTEGVTIGTSGLSGL	
IARHLVTEHGVRLDVLVSRQGPDPAGATELIAEQQLQAQVRVRSCTDTRAAALV	
DEVGPFAGVWAGALADATVDNLADALATLTAAKAHAAWMLHEHTQDHGLITLWVF	
SSLASVLGSGQAANYAANFLDALATRLTGLPTSLANGSWERTSGMSHLTRAD	
HDRLARAGLPLTDEQGLFDFGALVDLGQVTTAFDLPALGSGAVPVLRLGVR	
PRRAVAGADSGGAEPTERTATLAPDARERSVLTSVLEQVAAQLGHPDPGAPEDQAF	
KSIGFDSLTAVFERNLSMTGLQLPATLVPDYPAGAVTRYLLDQVSPAPAAAGPDD	
LDPERLURKRVFVRFVARFAGVLEALLKLAGDPDPARTPDABEPDEGLIDAMDTEA	
LIRHEUMDTGTA"	
9637..14925	

CDS

```
/note="orf2"
/codon_start=1
/transl_table=11
/product="PKS module 2"
/protein_id="AAC38062.1"
/db_xref="GI:2624949"
/translation="MTNEOLVRLAKVSAKENARLRRENASLOEGAAEPVAIVGMACRL
FGGVDSPESELWELMAGDASVPFPTDRCDWLDGLYDPPDHLGTAYVREGGFLDRAG
AFDPPFGISPREALAMDPOQRLLLETTWEVLERAGIDPTTLRGSRGTGFTGVLQDQY
VTRLDDPFAELGSLGNGSAASVASGRVSTFTFGEGPAVSLDTACSSSLVALHLVRP
ALRLLARLVAGTVVNSVPHYFVSFQRGLAFDARCKFSFASADGTVLSEGVGLLV
ERLSALRLVGVAVRGTAVNQDASNGLTAPTGPSQORVIRDALTDARLTADVDD
LLEAGHTGRLGDETAQALIATYGQDRDPVWLGSLKSNIGHTQAAAGVAGVITKVM
AMRGVMPRTLVSEPTPNVMSAGEVRLTWARDWPSDQRPRAAVSFGISGNTA
HVVLSCAAPAPDPVPVTPAGVPVWLSAHTGPAISALAEHLAHEDEAPLDGVA
HLENTFTRALPHRAVVVGTREEFVMAULRAVPTPRRGLPAVMFTGQSORPGMRQY
HYLTFVFAQLDEVPALATGCLREVMFDPQAEITLQRTDHAQIALFAPETALYRME
SWGLRPFMVCCHSVGEITAAHNGHATVLSGDRTLLHIAITQMLNTKPNLNVSHAFSP
HTLQPHLDNHDTTISAAINGPHATVLSGDRTLLHIAITQMLNTKPNLNVSHAFSP
MOPILQPTFTLNTLTHHPHTLISMLTATPHDPTHTWTOHITAPRYTDLHLHL
HNGITYLSEIGDPTTLALARTPTTLHLIPIITRNNEVSENEALGRVFSVGHVS
DNRALTGRRSLTTPPQRDQFMDHDAAGAEVGGAGTGDHLLGAVTTVADT
GELLISGRSTTHWLTHTVNGTVIVFETALIDLAHAAEHDTHTVDELIVTPTL
ALHTPTGTSVTGAETDCHRPALHSRTTGTWTRHTTGLSNTHOHPAAGLTSWSTP
DAHQIDLTAAYOQADTGLHYGPAFOGLHTLYQDNDTLPAEIELPEAAGLPTGHIHP
ALLDALHAPVSAARDEESGLOLPESWAGTVHSGATHLVRDLTDDTGALRLAT
DITGQPVITSLTTRPLHPDQHLHAAOPEINHLITLGNSTPTLPTTTCGHFERITYPT
AGDILLADTHRTIAQTLRIQOHLAADTTPLVVEATHDHLAGRVMLRTAQTETHP
GRILLDNDHTSDALASVASGKPOARIRDAITLRLTLRLSDVTRASTERRPDL
SMGVLVGTGGLSLIARHIVTEHGVLDLVLVSQGPADGATBLTNLOQOHGARI
RTISCDLTORTALTLLIDBIGPLTVGVVHTAGALADTTIDHLPDTLTTLTKANPAW
WHLHETQDHDALFVVFSSVAGVGNAGQANAAANSFLDALLTHERRQGLPETSIAW
GHWQREGHTAHLTQADHRMTRAGHGLTDAGTTLFTDLDRGGLGAVPVKHLHPT
LNRADTVAVLGLKRVAPRAAPQAPAESWADRIAALMPEERDRAVLDLVRQOSAR
VLGHSDADRIDPGQAFDGFSLTAVEFNRVTATGLRLPATLIFDPTFNALADH
LQAOLGPDGSGSSPEVVRDLERFERQLFDRPDDEVGTAVTAIVRLRSILAHLDAPAT
APAARTASAOVLDTLEDAQSSELLSFIDKEFGASN"
14964..21266
/notes="orf3"
/codon_start=1
/transl_table=11
/product="PKS module 3"
/protein_id="AAC38063.1"
/db_xref="GI:2624950"
/translation="MADSKTLDLYLKNVTADLYTRERLAEVESASAEPIAIGMACR
FPGVTSPEALMELVSSGDADLSGFPDDDRGDMWEGFFDSDDPAQEGTSYMRGEGFTNA
TGDFPAFFGISPREAVAMPDQORLLLETTWEVLERAGIDPTTLRGSRGTGFTGLMANS
YTRVVRIPGELESVAGNEAASVASGRVSTFTFGEGPAMSVDTACSSSLTALHLAVR
ALRSGECLPAGVGVTPATPRIFSDLSRQGLARDGRCARFSTADGTLSEGVGLL
LYVERLSAQLRGLSHVLGVVGTAVNQDASNGLTAPNGPAQORVIFAALADARLTAAD
IDLLIENHGTGKLDPIEAQALIAITYGQDRQEPQLWNGSLKSNIGHSQAAAGVAGIYKI
VMAHRGVMPKTLIHIDEPTPEVMSAGTVELLAEADWPVDPRPRRPAVSSFGVSG
NAHYLEQAPPAPEPTESAEVPESEAGFSGVLELGPSPVWVLSAQNETALKD
LARQATVDGSDVRIIAHSLISTRAALPHRAVVVGSSTREEFATALEQITGHTPDGKTV
LVPFGOSQWGMGQRLWTSPLYRDTIHEIDTALSEYIDWSDVTLTSGQAPTWDR
VDVLOPVLFTWMTALAQHGLTPDAIIGHSQGLIAAAYTAGALTQDAARLITLR
SKALLPLTGHGTMISVLAPQTIQDIAEDPHNTRIRWIAHNAFNPTITITGDTQALH
LQOOLSHHRIMRWQLPGVDFAGSHGIDQLHQLKILAPITPQKPTIFPFTSTNTW
IHTTLDADYVWNLRFQFOPAINHLNNGYTYTIESPHPLTSPFOETNPNNTT
ITTLHNSQNDTHALITAGHARTHGHPIWTWHTLIPPTKTPITPTPTFOETHWLNKTI
ADAVESAGLSRTHDPLIGATTLAHSGETVLNGLSLTHHGWIAOHVNGVPLPQGT
ALVDMALHAGDHTDTLLELIHTPTLTHTPTIQLTTPHTDNDTLTIYTRAP
EQPWIOHAGTLTRANTARSSPETIWPPTNAEPIIDGTYTQRIATLGLYGFAGQLH
ALMRDSDNRLKADGVLDPDITDNTSDCTDGYIHPALLDAAHLAVASDGEIRL
PFAWGTTLHATKARLVELDIDDDGTLSSATDTSQGVVTVDSITTRPIDPAQLR
APRQDLDYALMDLWPSGSAVAPAEFEQVEVPKGVDDVLADHTRITAOITLRVQOHL
AADTTPPLVVEATHDLAGAAVWGLRTAQTEHPGRIILIDTNDHPTSDALASVAS
GEQAEIRGALTLPLRTTAVTDALTTPPADADWHLDTSEGTLENLILITGSDTAIR
NLLPGEVRVAVRAAGLNFRDVLIALGKMYFGAETMGSEAGIVLIDIGDITLDPGRV
TGLFPNALANTAVADRMMTPIIPASWNSNTDAAFFIVILYATYGLHDLAALKEPGERVL
IHAGTGGQVTAALQLAQHGAEVAFATHTKMPHLRAMGLDHHIATSRTPIPTFFH
TTTQGGQGVHVLNLDLTLASLRLLTPNGRFIEMGKTVDVREAVTHPQVTVRA
FDIADDPDRIGAMVELTGLPERGALRGLPVRCDVRESVDAFRFMAQARHIGVV
```

Query Match 32.8%; Score 164.4; DB 1; Length 32870;  
Best Local Similarity 60.1%; Pred. No. 5.8e-09;  
Matches 273; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY	12	CGCCGAGCAGAACCGCATCCTGCTGAAGCTGTGTCGCGGCGCACGCTTCGACGGTGTCTCGG	71
Db	14415	CGAGGAGCGGCGACCGGCGGTGCTGACCTGTGTCGCGGCGCACGCTTCGACGGTGTCTCGG	14474
QY	72	CCACAGCGCGCGCGAAGGATCGGCGCGCGCGCACGCGGCGCACGCTTCGACGGTGTCTCGAC	131
Db	14475	GCATAGCAGCGCGCACCGGATCGACCGCGCGCGCGCACGCTTCGACGGTGTCTCGAC	14534
QY	132	GCTGCGCGCGGTCAACCTCCGCAACAGCCTGACCGCGCGCACGCGGCGCGGCTGCGGCTGCC	191
Db	14535	GCTCACCGCGGTGAGTTTCGACACCGCGCTCACCACCGCGCACGCGGCTGCGGCTGCC	14594
QY	192	GACGCTGATCTTCGACTACCCACCCCGGAGGCGCTGTGCTGCGTACCTGCGGCTCGA	251
Db	14595	CACCTGATCTTCGACACCCACCCACACCGAAGCGGCTGCGGCGATCATCTGCTGCCA	14654
QY	252	CCTGCGGAGGCGCGACGCGGCTGCGACGCGGCGGGAAGCAGCCTCCGGCGAGTCTCG	311
Db	14655	CGGCGCATTTGGGCGACGCGGCGGTGCTGCTCCCGGAGGCTGCTGCGGACCTGAGG	14714
QY	312	GGCGGTGCGGTTCCGCGCGGTTCAGAGGAGGCGGCGGTGCTGCGACACGCTGCTCGG	371
Db	14715	CGAACCGGCGATTTGTCGACACCGCGCGCGACGAGGTGGGCGCGGATCGCGTCCG	14774
QY	372	CGACACCGGCGACCGAACCGGCGCACCGACCGCGGAGACACCGAAGCGGCGCGCGCG	431
Db	14775	GTTGCGCTCCATCTCTGCGCCACCTCGACGCGCGCGCGCACCGGCGCGCGCGCGCG	14834
QY	432	CGACCGCAGACTGATCGACGCGCTGACATCTCC	465
Db	14835	CAGCGCCCGGCTCTCGACACGCTTGAGGAGCGCC	14868

RESULT 7  
BD217187  
LOCUS BD217187  
DEFINITION DNA encoding methymycin and pikromycin.  
ACCESSION BD217187  
VERSION BD217187.1 GI:33026957  
KEYWORDS JP 2002536959-A/17,  
Streptomyces venezuelae

ORGANISM Streptomyces venezuelae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 4689)  
AUTHORS Sherman, D.H., Liu, H.W., Xue, Y. and Zhao, L.  
TITLE DNA encoding methymycin and pikromycin  
JOURNAL Patent: JP 2002536959-A 17 05-NOV-2002;  
REGENTS OF THE UNIVERSITY OF MINNESOTA

COMMENT OS Streptomyces venezuelae  
PN JP 2002536959-A/17  
PD 05-NOV-2002  
PF 25-JUN-1998 JP 2000557373  
PR 26-JUN-1998 US 09/105537  
PI DAVID H. SHERMAN, HUNG WEN LIU, YONGQUAN XUE, LISHAN ZHAO PC  
C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/00, C12P7/ PC  
42, C12P19/62,  
PC C12N15/00, C12N5/00  
CC DNA encoding methymycin and pikromycin  
FH Key  
FT source 1. -4689  
FT Location/Qualifiers  
/organism='Streptomyces venezuelae'.  
/mol\_type='genomic DNA'  
/db\_xref='taxon:54571'

ORIGIN  
Query Match 31.7%; Score 158.6; DB 6; Length 4689;  
Best Local Similarity 59.0%; Pred. No. 4.5e-08;  
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;  
QY 1 GCGCTCCCGAGCGCGGAGCAACCGCATCTCTGTAAGCTGTCGCGGCGGCGTTCG 60  
Db 4183 GCGCTCCCGAGCGCGGAGCGCGGCGGCGTCTCTGTAAGCTGTCGCGGCGGCGTTCG 4242  
QY 61 ACGGTGCTCGCGGCGGCGGAGCGCATCTGCGCGCGCGCGGCGGCGGCGTTCGAGAGGTC 120  
Db 4243 GCGGTACTCGGCGCATCTCTCCCGAGCGGCGGCGCGGCGGCGGCGGCGTTCACGAGGTC 4302  
QY 121 GCGTTCGACTCGTGGCGCGGCGTCAACTCTCGCAACAGCTGCGCGCGGCGGCGGCGT 180  
Db 4303 GCGTTCGACTCGTGGCGCGGCGTCAACTCTCGCAACAGCTGCGCGGCGGCGGCGGCGT 4362  
QY 181 GCGGTGCGCGGCGGCGTCAACTCTCGCAACAGCTGCGCGGCGGCGGCGGCGTTCGCTG 240  
Db 4363 AGGCTCCCGCGGCGGCGTCTTGCACCAACCGCGCGGCGGCGGCGGCGGCGGCGGCG 4422  
QY 241 GCGGTGCGAACTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300  
Db 4423 CACGAGGCGTACTCGCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4482  
QY 301 CGAGTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 360  
Db 4483 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4542  
QY 361 CTCGCGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 417  
Db 4543 CTGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4602  
QY 418 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 477  
Db 4603 GACCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4662  
QY 478 CGAGCGCTCGGCG 490  
Db 4663 ATGGCTCTCGGCG 4675

RESULT 8  
AF079138  
LOCUS  
DEFINITION Streptomyces venezuelae methymycin/pikromycin linear BCT 30-OCT-1998

gene cluster, complete sequence.

AF079138  
VERSION AF079138.1 GI:3808326  
KEYWORDS Streptomyces venezuelae  
SOURCE Streptomyces venezuelae  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 37948)  
AUTHORS Xue, Y., Zhao, L., Liu, H.W. and Sherman, D.H.  
TITLE A gene cluster for macrolide antibiotic biosynthesis in  
Streptomyces venezuelae: architecture of metabolic diversity  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12111-12116 (1998)

COMMENT 98445333  
9770448  
2 (bases 1 to 37948)  
AUTHORS Xue, Y., Wilson, D., Zhuo, L. and Sherman, D.H.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-1998) Department of Microbiology, University of  
Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis,  
MN 55455, USA  
3 (bases 1 to 37948)  
AUTHORS Xue, Y., Wilson, D., Zhuo, L. and Sherman, D.H.  
TITLE Direct Submission  
JOURNAL Submitted (29-OCT-1998) Department of Microbiology, University of  
Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis,  
MN 55455, USA

REMARK  
Sequence update by submitter  
COMMENT On Oct 30, 1998 this sequence version replaced gi:3800831.  
FEATURES  
Location/Qualifiers  
1. 37948  
/organism='Streptomyces venezuelae'  
/mol\_type='genomic DNA'  
/strain='ATCC15439'  
/db\_xref='ATCC:15439'  
/db\_xref='taxon:54571'  
154. .1122  
/gene='pikR2'  
154. .1122  
/gene='pikR2'  
/note='mIs type resistance protein'  
/codon\_start=1  
/transl\_table=11  
/product='rRNA methyltransferase PikR2'  
/protein\_id='AAC69327.1'  
/db\_xref='GI:3800832'  
/translation='MAFSPGGRHELQNFIVDRSVIDEIDELVARTKGPILGPGD  
GALTPLSRGRPTAVELDGRRRAQRIGARTPGHVTVVHDDFLQYPLRPHPVHGVN  
PFHLTAIMRLDLDAQHHTAVLLVOWEVARREAGVGSTLLTAGAPWTFDLIRSV  
PARAFRPMGVGGVLAIRRSAPLVGVQVTKYQDFVRQVFTKGNGLKEILIRTRGIS  
ORDLATLWRNEISPHALPKDLKPGOWASLWELTGGTADGSDCTAGCGAGSHAAR  
VGAGHPGGRVSASRGVQPARRGHVAVRSTGTETPRWGRGAEASA'  
complement (1283. .2293)  
/gene='pikR1'  
complement (1283. .2293)  
/gene='pikR1'  
/note='mIs type resistance protein'  
/codon\_start=1  
/transl\_table=11  
/product='rRNA methyltransferase PikR1'  
/protein\_id='AAC69328.1'  
/db\_xref='GI:3800833'  
/translation='WAMRDSIPREADRTLRLRELQNFLODDRVRNLVTHVEGDGRN  
VLEIGPKGALTELVSTVYVEMDPHAAHVRKFEGERVTYVQGLDLEIRP  
DIDTVGNVPFGITQILRSLLSTNQSAALIVQWEVARREAGVGSTLLTAGAPWTFDLIRSV  
YEFVHVRVRSASSRPMFVRVDGGVLTIRRPQPLLPESASFAQNFQAEVFTGPRGL  
AEILRHIPKTYRLADRGIPDGLPKDITLQWALFQASQFSPAPGAPGTRMPG  
QGGGAGGRDYDSETSRRAVPGRRYGTRGGEPCAPRAQVETQTKGRQARGSSYGRRT  
GR'  
2912. .16753  
/gene='pikA1'  
2912. .16753  
/gene='pikA1'

```

/notes="multifunctional polyketide synthase; contains a
loading module, Pik module 1 and module 2"
/codon_start=1
/transl_table=11
/product="type 1 polyketide synthase PikAI"
/protein_id="AAC69329.1"
/db_xref="GI:3800834"
/transl_table="MSRAGITRTGARTPTVTCRGAADTCEVRVRRGLPAGPDHAEH
SFSRQAVTAEILIRGEMSTVSKSEEEFVSVDNAGSAGHTAEPAVAVGICRVPQ
ARDPREFWELAAAGQAVTDPVADRWAGDIPDRSPRASGRNMGFIEDVDRFDA
AFFGCSIPREAEVDPOORLALRGLWEALRAGIDPSLITGRTRGVFAGATWDDYATLK
HQCGAATTHITVTLGLRGTIANRLSYTLGRGPMVVDSCQSSSLVAHLACSILR
GESLALAGGVSLNIPDSIIIGASKFGCLSPDGRYTFDARANGVVRGEGGCVVLYR
LSRAVADGPDVLAIVRSGAAGQMTTPDAQAEVILREAHERRAGTAPADRVVY
ELHGTGTPGVDPIEAAALGAALGCRPAGQPLLGVSKTNIGHLEBRAGIAGLIKAVL
AVRGRAIPASUNYETNPDAIPFBEINLRNVTYILPWEHGDQRMVGVSVFGMGSTN
AHVLEAPGCGCRSTGVSTVGGSAVGGVPMVVSAAKSAALDAQIETLRAAASGRD
RTDGDVAGAVDAGAVAGAVARVLAGRAQPEHRAVVGSGPDDLAALAAPEGLVRG
VASGVAVFVPCGQOWACWGAELDDSSAVFAAAMAECEAALLSPYVDWLSLEAVRQ
APGAPTLERVDVQPTFVAVVSLARVWQHGVTPQAVVGHSGOGEIAAAVYAGALSLLD
DAARVTLRSKSIASHLAGKGMLSLASEDAVLERLAGPGLSVAAVNGTATVVSQ
DPVQIETELARACEADGVRAVPIVDYASHSGQVEIIESELAELVLAGUSPQAPRPFPS
TLEGATITPEVDGVTYRNLRHRVGFAPAVETLATDEGTFHFEVSAHPVLTWATERH
TVTGLATLRNDGQDLVSLAEAWANGLADWSPLLPSATGHSIDLPTVAFQTERH
WLGETALAPAGEACTSLAGVLRTEARPAELDRDEOLRVILDKVRAQTAQVLYGVATGG
QIEVDRTFREAGCTSLAGVLRNINRAAFGVMAPSMTDFPTPEALAEQILLVYHGE
AANAPGAPAPVAAAGAVDEPVALVGMACHRLPGVASPELMLRLVAGGDAIIEFPQ
DRMGDEGLHPDDEHPTGSVVRQGGFIENVAGFDAAFFGISPREALANDPQORILLE
TSMWEADAGIDPTSLRQGVYFTGAMTHEYGFSLRDEGLDGLYLLTGNTASVMSG
RVSVTLGLEGPALTVDYACSSLSVLAHLAVQALRKEVDMALAGVAMPTPGMFVEF
SRQGLAGDGRSKAFASADGTSWEGVLLVERLSDARNHGHOVLAVVRGSAALNQD
GASNGLTAPNGSQOVRTRALADARLTSDVDVVEAGHGTGLRGPDEAOLATYQ
QGRDDEPLRLGSKNIGHTQAAAGSVGVKIGMNAHVVLEAPVVEGASVVERSVG
AGAVELLTEADWPKQDGLRRAAVSFGISGNTAHVVLEAPVVEGASVVERSVG
GSAVGGVTPVVSAAKSAALDAQIETLRAAFASDRDTPDADAGAVAGAVHLADGR
AQFHRVALVAGADLLVQALADPDGLRTGASGVRAFPVPGQGTQWAGCAELDD
SSAVFAAAMAECEAALLSPYVDWLSLEAVRQAPGAPTLERVDVQPTFVAVVSLARV
QHGVTPQAVVGHSGOGEIAAAVAGALPDDARVTVTLRSKSIASHLAGKGMLSLAL
NEDAVRTPGQGLSVAAVNGTATVVSQDPVQIEELAQACKADGFRARIIPVDYAS
HSRQVEIIESELAQVLAGLSPOAPRVFPFTLEGTWIITEPVDIGTYTVRNLRHVGFA
PAIETVLLDEEGTFHFEVSAHPVLTWPTETVTLGLTRREGQGERLVTSLAEAWN
GLPVATWILLPATASRPGPLTVFAQEARWYLENTPAALATGDMRYIRDWKLPAAG
SERLGLSRLWATPDSHQAQAAVLTALVDAGAKVEVLTAGADDDRALAARLTALT
TGDGTGTVSYLLDGLVQVAVQALGDAGIKAPLWSVTCGASVGRLOTPTDGPAPRL
WGLGRVLEHPRWAGVLDLPQDPAALAHHLVTLASGATGEDQIARTPTGLHAPRL
ARAPHLRRPTLPATVPHGTFTVLITGGTALGSHAARMAHGHLLLSRSGEQAPGA
TQTLAELTAGARVTIAACDVADPHAMRTLLDAIPAEPLTAVVHTAGALDDGIVDTL
LAARRATGRSAVSVAWGPWGGGAAGDVAERLRNMGVPGMDPELAALESALGR
DETAITVADLDWRFLAYSGRPOPLVEELPEVERIIDARSAVSGGSSAQGANP
LAERLAAAPGERTELLGLVRAQAAVAVLRMSPEDVAADRPAKIDIGDSLAGVELRN
LRUTAGTQLPATVPHDPTPLAULSVLRSEPLFDEETADARSAALPATVVGAGAG
AGTDADDDP IATVAMSCRYPDIRSPEDILMRMLSEGGEGITFPPTDRGWLDGLYDAD
PDALGRAYVREGGFLHDAEFAEFVGSPREALANDPQORMLTTSWEAFERAGIEP
ASLRGSSTVTEGLSVQDYAAVPNAPRGVEGYLLTGPVSAQRG IATVFELEGPAT
TVDTACSSSLTALHLAVALRSSECTMALAGVAMWATPHMFVESRORALAPDGRSK
AFSADADGGAAGVGLLILVERLSDRNRNGHPVLAVRGTAVNDGASNLTPANGPS
QQRVTRQALADAPGDDIDA VETHTGTSLDGPIEAQQLQATYKREPRAPLAIGS
VAKSNIHTQAAAGAG IKNVLMHRHGTLPKTHADBPSPHYDWMANSGLALVTEPIDW
PAGTGPRRAAVSFGISGNTAHVVLENTPAALATGDMRYIRDWKLPAAG
VAEGSEASAPAPASGRREASLPGLHPWLVSAXDEQSLRQQAALHAWLSPEAADSIDA
DGPALRDVGYTLATSRTAFAHRAAVTAADRDGFLDGLATLAQGTSAHVHLDIARDQ
TTAFLFTGGQSPRGHRELYDRHPVFAHALDEI CHMLDGLHLEPLLDVWFAEGSAAE
AALLDETRYTQCALFALEVALFVLSWGRNPPAALLGHSVGSVIAAHAVNGVSLADAA
RILVAARQLMQELPAGMALVQAADERIVMLETEERYAGRLDVAAVNGPEAAVLSG
DADAAREAYVSGLGRTRALVRSHAFSAHMDGMDFRAVLETFEVRPRLSTVVS
NVTGLAGPDDLCDPEYVHVHVGTVRFLDGVRLDGLVTRCTRLGPDGLVTAMAAD
GLADTPADSAAASPVGSPAGSDSAGALRPRPLVALLRRKRSETEYVADLGPRAH
AHGTGPDHMAVAGSAHRVLDLPTYSFRDRYVLDPAADAVDTAGLGLGTADHLLV
GAVSLIPDRDLGLLTLRGLSHRTHADHVLGSLVLLPGAAWVELAAHAAESAGRDV
RELTLLEPLVLPHGGVGLRVTVWAPAGEPGESAGDGRPVSLHNGSLRADAPAGTAW
CHATGLTRPELVPDRAAMVPPQGAEEVFLDGLYERLDNGSLAFGLFGLNLA
WRYEVEVFIADIALPATNATATANGGGSAAAAPYGIHPALLDASLHAIANGVELDE

```

```

PELVRFPHWSGVTVHAAGAAAARVRLASAGTDAVSLTIDGEGRPLVSVRLRLRPV
TADQAAASRVGGLMHRVAMPYPYALASSGEQDPHATSYGPTAVLKGKELKVAALLESAG
VEYGLXPDLAALSQDVAAGAPARTVTLAPLIPAGDAGGAEGLVTRTAQTNPGRFLGLDLA
DEHGLTLLLVTRGAVRDPGSGADGDGDELSHAAAGLVRTAQTNPGRFGLDLDA
DDASXYTLSPSLSDAGIRDEPQALHDGTIRLARLASVRETCCTAAPALAPAEQTVLL
TGGTGGGLGVHVVHVGWGVRRLLVSRGTDAPAGDELVHELEALGADVSAACDV
ADREALTAVLDAIPAEHLTAVVHTAGVLSGDTIPSMTTEDEHVLRPKVDAFLIDE
LWSTPMDLAEAFVNFSSAAAVFGAGQCAAYAAANATLDAALRRRAAGLPALSIGWGL
KAEYSGMTQGLQADLRMRSGRIGGISDAEGIALDDALDRHPVLLPRLDAAGL
RDAGNDPAGIPALFRVDVVGARTVRAVPSAASASTTAGTCTGTGAAGTAATAVNL
DRAATVGPAPKORILLLEFVGEVAVLHARGHRIIDAERGFLDGFOSLTAVELNRLL
NSAGLGLPATKILVDFHPSPALASHLDAERHGRASDQDCAGRNNGENGTASSTAE
TDALLAQTRLLEGALVLTGSDAPGSEVEVLEHLKSLRSMVTGETGTGTASGAPDAGS
GAERDPAWAGDAGGSGEDGAGVDPFMNASAEELFGLDDQDPSD"
16858..28077
/gene="PikAI"
16858..28077
/gene="PikAI"
/notes="multifunctional polyketide synthase; contains Pik
module 3 and module 4"
/codon_start=1
/transl_table=11
/product="type 1 polyketide synthase PikAI"
/protein_id="AAC69330.1"
/db_xref="GI:3800835"
/transl_table="MSTVNEEKYLDYLRRTADLHEARGLRELEKAKAGEPVAVGMA
CRPGVASPEDLRLVAGGDAISEFPQDRGMDVEGLYDPNPATGKSAREAGFLY
EALGDDDFGFSIPREALANDPQORILLLEASWEAFHAGIPAAATAGTSVGVFTGMV
HDYATRLTDVEGIEGLTGNSSVASGRVAYTLGLRPAVTVDVTCSSLSVALHLA
VQALRKSEVDNALAGGVTVMTSPSTFVEFSGRGLAPNGRSKSFSTADGTWSNBSVG
VLLVLRSDARRKGRHILAVRGTAVNDQSGSLTAPGRSQVIRRALADAKLT
SDVVEAHRHGVLPKTLHEKPTIDQVMSAGAVELLTEAMDPWDGGLRRAVVSF
VIMVQMRHGVLPKTLHEKPTIDQVMSAGAVELLTEAMDPWDGGLRRAVVSF
GVSTNAHVLEAPAEETPASPATPAVPSVGLVPLVSAKTPAALDAQIGRLA
AFASQGRDADPGAVARVLAGGRAEFHRAVVLGTQDDPFAQALTAPEGLRGTPSD
VGRVAVFPQGQTMAGMGAELDVSKFEAAMACEALSALSRVYDWSLEAVVRQPGA
PTLERVDVQPTFVAVVSLARVWQHGVTPQAVVGHSGOGEIAAAVNGTATVVSQDTPQ
VTVTLRSKSIASHLAGKGMISLASEETATROIRIENHGLSIAAVNGTATVVSQDTPQ
IQELAAQACEADGVRAVPIVDYASHSGQVEIIESELAELVLAGUSPQAPRPFPS
WITTEPDLGTYTVRNLRHVGFAVETLATDEGTFHFEVSAHPVLTWPTETVGTG
AGTLRREGQGERLVTSLAEAWNGLTIDWAPVLTATGHEPTELYAFORSHYMLHD
SPAVGVSQDSDWRVIRIDWKRLAASERAGUSGRWLVVVPEDRGAERAPVLAALSQA
GADPQLVDSPGLDRQRLAATLGEALAAAGVADVGLSLLADESAHPCHPAPFTRGT
GATLTVALDEADGAAVAPLWCVHGASVGRADHVTSPQAQWVMGRVAALHPERW
GSLIDLPADADRALDRMTTVLAGTGDQDOVAVRASGLIARLRLVRAISPAHGTASPMW
QADGTVLTGAEPEAPAAEARLARDGAGHLLHTTTPSGEAGETSGAAEDSGLAGL
VAELADGATATVTCDLTDABAARLLAGVSDAHLPLSNVHLPTVDSSEPLAATDAD
ALARVVVTAALHDLRLLEAAAGRPVVLVLFSSVAALWGAAGQCAAGATFL
DALAQHRAKADGPTVTSVAMSWMGSRVTEGATGERLRRLRGLRPLAPATALTALDIALG
HGDATVITADVWMSFAPGFTTARPGTLLADLPEARALDEQOSTTAADDTVLSRELG
ALTGAEQQRMOELVREHLAVVNLHPSPEAVDTGRAFDLGFDSLTAVELRNLKNAT
GLALPATLVDFYPTPTLAEFLAELIHLGEQAGAGQLPVDGGVDDEPVAIVGMACRLP

```

Query Match	31.7%	Score	158.6	DB 1	Length	37948	
Best Local Similarity	59.0%	Pred. No.	2.2e-08				
Matches	291	Conservative	0	Mismatches	199	Indels	3
Gaps	1						
QY	1	GGGTCCTCCGACCGCGAGCAGAACCGCATCTCTGCTGTAACCTGCTCCGCGCGCCACGCTTG	60				
DB	32343	GGGTCCTCCGACCGCGCGCGCGCGCGCTCTCTCACCTCGTCCGTACCCACCGCGCG	32402				
QY	61	ACGGTGTCTCGGCACACGCGCGCGCGCGCATCGCGCGCGCGCGCGCGCTTCCAGGAGGTC	120				
DB	32403	GGGTACTCGGCATCTCTCCCGGACCGGGTGGCGCGCGCGCTCTTCCACCGAGCTC	32462				
QY	121	GGCTTCGACTCGCTGCCCGCGGTCAACCTCCGCAACAGCCTGCACCGCGCCACCGCGCTG	180				
DB	32463	GGCTTCGACTCGCTGCAGCGCGCGGTGAGCTCCGCAACACAGCTCTTCCACGGTGGTGGCGAAC	32522				
QY	181	CGGCTCCCGCGACGCTGATCTTCGACTACCCACCGCGAGCGCTGCTCGGCTACCTG	240				
DB	32523	AGGCTCCCGCGCACCAACGCTCTTCGACCAACCGCGCGCGCTCTCGCGCGCACCTC	32582				









Db 29735 GACCTGTGGAGCGGAGCGTGCATGACCACTGGAGCGGAGCCCTGATCGG 29794  
Qy 478 CGAGCCCTCGGC 490  
Db 29795 ATGGCTCTCGGCC 29807

RESULT 13  
AX697993  
LOCUS AX697993 4725 bp DNA linear PAT 02-APR-2003  
DEFINITION Sequence 17 from Patent WO03010193.  
ACCESSION AX697993  
VERSION AX697993.1 GI:29499042  
KEYWORDS  
SOURCE Micromonospora carbonacea

ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE 1  
AUTHORS Farnet, C.M., Staffa, A. and Yang, X.  
TITLE Genes and proteins for the biosynthesis of rosaramicin  
JOURNAL Patent: WO 03010193-A 17 06-FEB-2003;  
Ecopia Biosciences Inc. (CA)

FEATURES  
source  
1. 4725  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

## ORIGIN

Query Match 31.1%; Score 155.8; DB 6; Length 4725;  
Best Local Similarity 62.2%; Pred. No. 8.7e-08;  
Matches 263; Conservative 0; Mismatches 157; Indels 3; Gaps 1;

Qy 12 GCGGAGCAGAACCGCATCTCTGCTGAAGCTGGTCCGGGCGACGCTTCGACGGTCTCGG 71  
Db 4254 CGACGAAACCGCGGGGCTCTCGACCTGTAGCGCGGAGGTTCGGCGCTCTCGG 4313

Qy 72 CCACAGCGGCGCGAGGATCGGCGCGCGCGCGCTTCAGGAGGTTCGACTTCGATC 131  
Db 4314 CCACCGGGGCGCGAGCAGCTCGGCGCGCGCGCTTCGGGAGATCGGATTCGATC 4373

Qy 132 GCTGCGCGCGGTTCACCTTCGCGCAACAGCTGCACGCGCGCGCGCGCTTCGCGCTCGC 191  
Db 4374 GCTGACCGCGCTGCGACCTGCGCGCGCGCGCTTCAGGCGCGCGCTTCGCGCTCGC 4433

Qy 192 GACGCTGATCTTCGACTACCCACCGCGAGCGCTGGTTCGCGCTACCTTCGCGTTCGAACT 251  
Db 4434 CACCTCTGCTTCGACCAACCCACCGCGAGCGGTTCGCGCGCGCGCTTCGCGCGCTGCT 4493

Qy 252 CTTGGGAGCGCGAGCGGCTTCGAGCGGCGGCGAGACGACCTTCGCGCGAGTCTTCGC 311  
Db 4494 CGGTCCCG 4550

Qy 312 GCGCGTGCCTTCG 371  
Db 4551 CGACTTCGCGTTCG 4610

Qy 372 CGACACCGGCGACCGAACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 431  
Db 4611 GGGGCTGAGCGCGAGCGGCTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4670

Qy 432 CGA 434

Db 4671 CGA 4673

RESULT 14  
AX697977  
LOCUS AX697977 60196 bp DNA linear PAT 02-APR-2003  
DEFINITION Sequence 1 from Patent WO03010193.  
ACCESSION AX697977

VERSION AX697977.1 GI:29499034  
KEYWORDS Micromonospora carbonacea  
SOURCE Micromonospora carbonacea  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.  
REFERENCE 1  
AUTHORS Farnet, C.M., Staffa, A. and Yang, X.  
TITLE Genes and proteins for the biosynthesis of rosaramicin  
JOURNAL Patent: WO 03010193-A 1 06-FEB-2003;  
Ecopia Biosciences Inc. (CA)  
FEATURES  
source  
1. 60196  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

## ORIGIN

Query Match 31.1%; Score 155.8; DB 6; Length 60196;  
Best Local Similarity 62.2%; Pred. No. 3.6e-08;  
Matches 263; Conservative 0; Mismatches 157; Indels 3; Gaps 1;

Qy 12 GCGGAGCAGAACCGCATCTCTGCTGAAGCTGGTCCGGGCGACGCTTCGACGGTCTCGG 71  
Db 40545 CGACGAAACCGCGGGGCTCTCGACCTGTACGCGCGGCGAGGTTCGCGCGCTCTCGG 40604

Qy 72 CCACAGCGGCGCGAAGGCGATCGGCGCGCGCGCGCTTCGAGAGGTTCGCTTCGATC 131  
Db 40605 CCACCGGGGCGCGAGCAGCTCGGCGCGCGCGCTTCGCGGAGATCGGATTCGATC 40664

Qy 132 GCTGCGCGCGGTTCACCTTCGCGCAACAGCTGCACGCGCGCGCGCGCTTCGCGCTCGC 191  
Db 40665 GCTGACCGCGCTTCGACCTTCGCGCAAGCGCTTCAGGCGCGCGCTTCGCGCTTCGCG 40724

Qy 192 GACGCTGATCTTCGACTACCCACCGCGAGCGCTGGTTCGCGCTACCTTCGCGCTTCGAACT 251  
Db 40725 CACCTCTGCTTCGACCAACCCACCGCGAGCGGTTCGCGCGCGCGCTTCGCGCGCTGCT 40784

Qy 252 CTTGGGAGCGCGACGACGCGCTTCGAGCGGCGGCGAGACGACCTTCGCGCGAGTCTTCGC 311  
Db 40785 CGGTCCCG 40841

Qy 312 GCGCGTGCCTTCGCGCGCGGTTCGAGGAGCGGCGCGGTTCGAGACGCTTCGCGCTTCGC 371  
Db 40842 CGACTTCGCGTTCGCGCGCGGTTCGCGGCGCGCGCTTCGAGCGCGCTTCGCGCTTCGC 40901

Qy 372 CGACACCGGCGACCGAACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 431  
Db 40902 GGGGCTGAGCGCGAGCGGTTCGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCG 40961

Qy 432 CGA 434

Db 40962 CGA 40964

RESULT 15  
AX607061/c  
LOCUS AX607061 15970 bp DNA linear BCT 27-APR-2001  
DEFINITION Mycobacterium tuberculosis CDC1551, section 147 of the complete genome.  
ACCESSION AE007061 AE000516  
VERSION AE007061.1 GI:13881778

## KEYWORDS

SOURCE Mycobacterium tuberculosis CDC1551

## ORGANISM

Myobacterium tuberculosis CDC1551  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.  
REFERENCE 1 (bases 1 to 15970)

## AUTHORS

Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Hart, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,



Best Local Similarity 65.4%; Pred. No. 7.9e-08;  
Matches 246; Conservative 0; Mismatches 121; Indels 9; Gaps 1;

QY	5	TCCCGGAGCGGAGCAGACCGCATCCTGCTGAAGCTGGTCCGCGGCCACGCTTTCGACGG	64
Db	3825	TGCCCCGAAGACGAGCAACACGCGCTGCTGCTGACCTGGTGGCTCGCACATCGGCCACCG	3766
QY	65	TGCTCGGCCACACGCGCGCCGGAAGCATCGGCGCCGCGCCAGGCGTTTCCAGGAGGTGGGT	124
Db	3765	TGCTGGGTAGCCAGCCAGCCCCGAGGCCATCGACCCGGATAGGGCATTCAGGAGTTGGGT	3706
QY	125	TGGAATCGGTGGCGCGGCTCAACCTCCGCAACAGGCTGCAACGCGGCCACCGGCTGCGGC	184
Db	3705	TGGAATCGGTCAACCGGCTCGAATCGCAACCGGCTCAAAATCGGCCACCGGCTGGGC	3646
QY	185	TGCCCGGAGCGGTGATCTTGACTACCCACCGCGGAGCGCTGGTCCGCTACCTGCGCG	244
Db	3645	TTTACCCACGCTCATCTTGACTACCCCACTCCGCGCGCTGGCGGCTATATGCGTC	3586
QY	245	TGGAATCTCT-----GCGGAGGCGCGACGCGCTGGACGCGCGCGGGAAGACGACC	295
Db	3585	GAGAACTGCTCGGCTCATCACCGCAAGACACTTCAGCGTGGCGCGCGGGAAGCCGAC	3526
QY	296	TCCGCGGAGTCTTCGCGGCGGTGCGCTTCGCGCGGTTCAGGAGGCGGCGCTGCTGGACA	355
Db	3525	TACAAACGCAATTGTGGCGTCCATTCCGGTCAAGCGCTTACGCGAGGCGGAGTGTGATC	3466
QY	356	CGCTGCTCGGCGCTCGC	371
Db	3465	TGTGCTCGGCTGGC	3450

Search completed: June 17, 2004, 01:40:46  
Job time : 2280.33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 22:08:13 ; Search time 285 Seconds  
(without alignments)  
7467.885 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543

Perfect score: 501  
Sequence: 1 ggggtcccgagcgagca.....cctcggcgagcagctga 501

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001as:\*
- 5: Geneseq2001bs:\*
- 6: Geneseq2002s:\*
- 7: Geneseq2003as:\*
- 8: Geneseq2003bs:\*
- 9: Geneseq2003cs:\*
- 10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	501	100.0	53789	2 AAV21187	AAV21187 Amycolato
2	158.6	31.7	4689	3 AAZ87299	Aaz87299 S. venez
3	158.6	31.7	36778	3 AAZ87318	Aaz87318 S. venez
4	158.6	31.7	37948	3 AAZ87285	Aaz87285 S. venez
5	158.6	31.7	38506	3 AAAY5633	Aay5633 Nucleotid
6	158.6	31.7	38506	3 AAZ56001	Aaz56001 Recombina
7	158.6	31.7	38506	7 ABE56090	Abe56090 S. venez
8	158.6	31.7	38506	7 ADA09418	Ada09418 Cosmid pk
9	155.8	31.1	4725	7 AAD55818	Aad55818 Micromono
10	155.8	31.1	60196	7 AAD55810	Aad55810 Micromono
11	154.4	30.8	110000	4 AA199682	AA199682 22
12	154.4	30.8	110000	4 AA199683	AA199683 22
13	152.6	30.5	65140	4 AAD17184	Continuation (23 o
14	152.6	30.5	125401	4 AAD17186	Aad17184 Streptomy
15	148.8	29.7	11238	7 AAD55817	Aad55817 Micromono
16	144.8	28.9	113193	7 AAD55815	Aad55815 Streptomy
17	142	28.3	27541	4 AAD17185	Aad17185 Streptomy
18	141.2	28.2	2700	4 AAD17183	Aad17183 ERD48 ins
19	140.2	28.0	11220	3 AAZ87298	Aaz87298 S. venez
20	139.6	27.9	30690	3 AAAY9301	Aay9301 S. avermi
21	139.6	27.9	30690	4 AAH79277	Aah79277 Streptomy
22	135.8	27.1	43280	2 AAH80413	Aah80413 Ty lactone
23	134.2	26.8	3072	7 ACA37796	Aca37796 Prokaryot

24	133.6	26.7	5676	2 AAV21186	AAV21186 Amycolato
25	132.4	26.4	103599	4 ABX04971	Abx04971 S. cinnam
26	132	26.3	44377	2 AAT80414	Aat80414 Platenoli
27	132	26.3	44377	2 AAT78508	Aat78508 Platenoli
28	130.8	26.1	29879	2 AAQ46806	Aaq46806 eryA regi
29	130.8	26.1	103599	4 ABX04971	Abx04971 S. cinnam
30	127.4	25.4	50937	3 AAA09469	Aaa09469 Streptoco
31	126.2	25.2	561	4 AAA89135	Aaa89135 Polyketid
32	126	25.1	6360	7 ACC42735	Acc42735 Geldanamy
33	125.6	25.1	47981	4 AAF30757	Aaf30757 Micromono
34	124.8	24.9	1681	6 AAS18438	Aas18438 Contig 11
35	124.4	24.8	77536	3 AAA14651	Aaa14651 Nucleotid
36	124	24.8	31422	3 AAA92302	Aaa92302 S. avermi
37	124	24.8	31422	4 AAH79278	Aah79278 Streptomy
38	123	24.6	82746	7 AAL61224	Aal61224 Actinosyn
39	122	24.4	20394	4 AAF24892	Aaf24892 Pimaricin
40	122	24.4	77536	3 AAA14651	Aaa14651 Nucleotid
41	120.6	24.1	75236	7 ABV75557	Abv75557 Saccharop
42	119.4	23.8	14055	7 AAL61170	Aal61170 Actinosyn
43	119.2	23.8	6210	7 ACC42734	Acc42734 Geldanamy
44	118.4	23.6	4466	3 AAA14663	Aaa14663 Nucleotid
45	118.4	23.6	4478	3 AAA14661	Aaa14661 Nucleotid

## ALIGNMENTS

RESULT 1	
AAV21187	
ID	AAV21187 standard; DNA; 53789 BP.
AC	
XX	
XX	AAV21187;
DT	24-JUL-1998 (first entry)
XX	
DE	Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX	
KW	Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW	polyketide synthase; actinomycete; ansamycin; ds.
XX	
XX	Amycolatopsis mediterranei.
OS	
XX	
Key	Location/Qualifiers
FT	CDS
FT	1825..15543
FT	/*tag= a
FT	/label= ORF_A
FT	/product= "polyketide synthase"
FT	15550..30759
FT	/*tag= b
FT	/label= ORF_B
FT	/product= "polyketide synthase"
FT	30895..36060
FT	/*tag= c
FT	/label= ORF_C
FT	/product= "polyketide synthase"
FT	36259..41325
FT	/*tag= d
FT	/label= ORF_D
FT	/product= "polyketide synthase"
FT	41373..51614
FT	/*tag= e
FT	/label= ORF_E
FT	/product= "polyketide synthase"
FT	51713..5293
FT	/*tag= f
FT	/label= ORF_F
FT	/product= "polyketide synthase"
XX	
PN	WO9807868-A1.
XX	
PD	26-FEB-1998.
XX	
PP	18-AUG-1997; 97WO-EP004495.

```
XX PR 20-AUG-1996; 96EP-00810551.
XX PA (NOVS ) NOVARTIS AG.
XX PI Schupp T, Toupet C, Engel N;
XX DR WPI; 1998-169172/15.
XX DR P-PSDB; AAW52845, AAW52846, AAW52847, AAW52848, AAW52849, AAW52850.
XX PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used to
XX PT produce rifamycin and rifamycin analogues.
XX PS Claim 4; Page 53-102; 205pp; English.
XX CC The present sequence represents a Amycolatopsis mediterranei rifamycin
XX CC synthesis gene cluster DNA fragment from the present invention. The DNA
XX CC fragment comprises a DNA region involved directly or indirectly in the
XX CC gene cluster responsible for rifamycin synthesis, including the adjacent
XX CC DNA regions to the right and left which, by reason of their function in
XX CC connection with rifamycin biosynthesis, qualify as constituents of this
XX CC rifamycin gene cluster, and functional fragments, derivatives or
XX CC constituents of these. The Amycolatopsis mediterranei rifamycin synthesis
XX CC gene cluster DNA fragment can be used for producing rifamycin, rifamycin
XX CC analogues or precursors. It can also be used for inactivating or
XX CC modifying genes involved in anamycin or rifamycin biosynthesis. The DNA
XX CC can be used for constructing mutant actinomycetes strains from which the
XX CC natural rifamycin or anamycin biosynthesis gene cluster has been partly
XX CC or completely deleted. The DNA fragment can be used for assembling a
XX CC library of polyketide synthases, which can be used for assembling a
XX CC library of polyketides. A hybridisation probe of the invention can be
XX CC used for identifying DNA fragments involved in the biosynthesis of
XX CC anamycins
XX SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 U; 0 Other;

Query Match 100.0%; Score 501; DB 2; Length 53789;
Best Local Similarity 100.0%; Pred. No. 4.8e-71;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGTCCCGCAGCGAGCAGAACCGATCTCTGCTGAAGCTGTCGCGCGGCCACGCTTCG 60
DB 15043 GCGGTCCCGCAGCGAGCAGAACCGATCTCTGCTGAAGCTGTCGCGCGGCCACGCTTCG 15102
QY 61 ACGGTGCTCGGCCACAGCGGCGCGAAGGCATTCGGCCGCGCGAGCGCTTCAGAGAGTTC 120
DB 15103 ACGGTGCTCGGCCACAGCGGCGCGAAGGCATTCGGCCGCGCGAGCGCTTCAGAGAGTTC 15162
QY 121 GCTTTCGACTGCTGGCGCGGTCAACCTTCGCAACAGCTTCGACGGGCCACCGGGCTG 180
DB 15163 GCTTTCGACTGCTGGCGCGGTCAACCTTCGCAACAGCTTCGACGGGCCACCGGGCTG 15222
QY 181 CGGCTGCGCGGACGCTGATCTTCGACTACCCACCCGAGGCGCTGGTGGCTACCTG 240
DB 15223 CGGCTGCGCGGACGCTGATCTTCGACTACCCACCCGAGGCGCTGGTGGCTACCTG 15282
QY 241 CGCGTCGAATCTCTCGGGAGGCGGACGACGACGCTTCGAGCGGGGAGACACCTCCGG 300
DB 15283 CGCGTCGAATCTCTCGGGAGGCGGACGACGACGCTTCGAGCGGGGAGACACCTCCGG 15342
QY 301 CGAGTCTCTCGGGCGCTGCTGGTTCGCGGTTCAAGAGGCGGGCGTGTGACACGCTG 360
DB 15343 CGAGTCTCTCGGGCGCTGCTGGTTCGCGGTTCAAGAGGCGGGCGTGTGACACGCTG 15402
QY 361 CTCGGCTTCGCGGACACCGGACCGGACCGGACGCGCGGACGACACCGAGCGGC 420
DB 15403 CTCGGCTTCGCGGACACCGGACCGGACCGGACGCGCGGACGACACCGAGCGGC 15462
QY 421 CGCGCGCGCGGACGAGCGAGAACTGATTCGACGACCTGGACATCTCCGGTCTCTGTGCAACGA 480
DB 15463 CGCGCGCGCGGACGAGCGAGAACTGATTCGACGACCTGGACATCTCCGGTCTCTGTGCAACGA 15522
QY 481 GCCCTCGGCGACGAGCTGA 501
```

```
DB 15523 GCCCTCGGCGACGAGCTGA 15543

RESULT 2
AAZ87299
ID AAZ87299 standard; DNA; 4689 BP.
XX AC AAZ87299;
XX DT 15-SEP-2003 (revised)
XX DT 05-JUN-2000 (first entry)
XX DE S. venezuelae macrolide biosynthetic gene pikAIII, SEQ ID NO:34.
XX KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
XX KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
XX KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
XX KW chronic obstructive pulmonary disease; respiratory inflammation;
XX KW hypercholesterolaemia; crop protection agent; ds.
XX OS Streptomyces venezuelae; ATCC15439.
XX FH Key Location/Qualifiers
XX FT CDS 1..4689
XX FT /*tag= a
XX FT /product= "PikAIII"
XX PN WO200000620-A2.
XX PD 06-JAN-2000.
XX PF 25-JUN-1999; 99WO-US014398.
XX PR 26-JUN-1998; 98US-00105537.
XX (MINU ) UNIV MINNESOTA.
XX Sherman DH, Liu H, Xue Y, Zhao L;
XX WPI; 2000-160679/14.
XX P-PSDB; AAY77194.
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX synthesis of methymycin and pikromycin.
XX Claim 15; Page 415-417; 438pp; English.
XX The invention relates to an isolated and purified nucleic acid segment
XX comprising a desosamine biosynthetic gene cluster, a fragment or its
XX biologically active variant, where the nucleic acid sequence is not
XX derived from the eryC gene cluster of Saccharopolyspora erythraea or
XX Streptomyces antibioticus. The invention also relates to a macrolide
XX biosynthetic gene cluster, or fragments thereof. The macrolide
XX biosynthetic gene cluster encodes proteins which synthesise methymycin,
XX pikromycin, neomethymycin, narbomycin or a combination of these
XX compounds. Recombinant or augmented cells comprising the desosamine
XX and/or macrolide biosynthetic gene clusters are useful for the production
XX of biologically active macrolides. The macrolide biosynthetic proteins
XX are useful for synthesis of methymycin, pikromycin, neomethymycin and
XX narbomycin. The alternative termination of polyketide synthesis may be
XX useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
XX monomers. The compounds produced by the recombinant host cells are useful
XX as biopolymers, e.g., in packaging or biomedical applications, to
XX engineer PHA monomer synthases or to prepare biologically active agents,
XX such as chemotherapeutics, immunosuppressants, agents to treat asthma,
XX chronic obstructive pulmonary disease as well as other diseases involving
XX respiratory inflammation, cholesterol-lowering agents or macrolide-based
XX antibiotics which are active against a variety of organisms, e.g.,
XX bacteria, including multi-drug resistant pneumococci and other
XX respiratory pathogens, as well as viral parasitic pathogens, or as crop
XX protection agents (e.g., fungicides or insecticides) via expression of
XX polyketides in plants. Sequences AAZ87295-287302 represent macrolide
```

CC	biosynthetic genes from Streptomyces venezuelae ATCC 15439, which encode
CC	proteins AAY7190-Y77197. (Updated on 15-SEP-2003 to standardise OS
CC	field)
XX	
SQ	Sequence 4689 BP; 648 A; 1882 C; 1572 G; 587 T; 0 U; 0 Other;
	Query Match 31.7%; Score 158.6; DB 3; Length 4689;
	Best Local Similarity 59.0%; Pred. No. 7.6e-17;
	Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;
QY	1 GCAGTCTCCCGACGCCGAGCAGAACCGCATCTGTGTGAAGCTGGTCGCGGCACGCTTCG 60
DB	4183 GCGCTCCCCGAGCCCGAGCGCCGCGCGCTCTCACTCCCTCFGCTCCATCCACACGCGCG 4242
QY	61 ACAGTGCTTCGGCCACACAGCGCGCCGAGGAGCATCGGCCGCGCCACGAGGGTTCCACGAGATC 120
DB	4243 GCCGTACTTCGGCATTTCTCCCCGACCGGGTGGCCCCCGCGCTGCTTACCAGAGCTC 4302
QY	121 GGCTTGACTCGTGTGGCGGGGTCAAACCTTCGCGAAGCTGTGACGCGCACCGGGCTG 180
DB	4303 GGGTTTGACTCGTGTAGCGCGCGTGCAGTTCGGCAACACAGCTCTCCACGGTGTGCGCAAC 4362
QY	181 CGGCTGCCCGGAGCGTGTGACTTCGACTACCCCAACCGCGAGGCGCTGGTTCGGTACTCGT 240
DB	4363 AGGCTCCCCGCGCACACAGGTTCTTGACACACCGACGCGCCGCGACCTCGCGCGCACCTC 4422
QY	241 CGGCTCGAACTCTGTGCGGGAGCGCAGACGCGCTTGACGCGGGCGGGAAGACGACCTCCGG 300
DB	4423 CACGAGGGGTACTTCGACCGCGCGAGCGCGCCCAGAGACTGGAGGGCGGGTGGC 4482
QY	301 CGAGTCTTCGGCGCGTGTGCGCTTCGCGCGTTCAAGAGGCGGCGCTGCTGGACACGCTG 360
DB	4483 CGGGCCCTTGCGCGAATCGCCCTCGACCGGTGCGGGACGCGGGGCTCTCGACACCGCTC 4542
QY	361 CTCGGCTCGCCGACACGCGCACCGACCGAACCGGCGACGAGCG---CCGAGACACCGAGAAG 417
DB	4543 CTGCGCTCTACCGGCACTCGAGCCCGAGCCGGGTTTCGCGCGGTTCCGACGCGGCGCGCC 4602
QY	418 GCCCGCGCGCCGACGACGAGAACTGATCGACGACTGGACATCTCCGGTCTCGTGCAA 477
DB	4603 GACCTGGTGGGAGCGGAGGGCTCGATCGACGACTTGACGCCGAGGCCCTGATCCGG 4662
QY	478 CGAGCCCTCGGGC 490
DB	4663 ATGSGCTCTCGGCC 4675
RESULT 3	
AAZ87318	
ID	AAZ87318 standard; DNA; 36778 BP.
XX	
AC	AAZ87318;
XX	
DT	15-SEP-2003 (revised)
DT	05-JUN-2000 (first entry)
XX	
DE	S. venezuelae pik (macrolide biosynthesis) gene cluster.
XX	
KW	Desosamine biosynthesis; macrolide; polyketide; polymycin; pikromycin;
KW	neomethymycin; naribomycin; polyhydroxyalkanoate monomer synthase;
KW	biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW	chronic obstructive pulmonary disease; respiratory inflammation;
KW	hypercholesterolaemia; crop protection agent; ds.
XX	
OS	Streptomyces venezuelae; ATCC15439.
XX	
FH	Key Location/Qualifiers
CDS	1742..15583
FT	/tag= a
FT	/product= "Pik gene cluster protein #1 (AAY77200)"
FT	15888..26907
FT	/tag= b
FT	/product= "Pik gene cluster protein #2 (AAY77201)"
FT	



[illegible]

RESULT 4	
AAZ87285	
ID	AAZ87285 standard; DNA; 37948 BP.
XX	
XX	
AC	AAZ87285;
XX	
DT	15-SEP-2003 (revised)
DT	05-JUN-2000 (first entry)
XX	
XX	
DE	S. venezuelae pik (macrolide biosynthesis) gene cluster, SEQ ID NO:5.
DE	
XX	
KW	Desoamine biosynthesis; macrolide; polyketide; polymycin; pikromycin;
KW	neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW	biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW	chronic obstructive pulmonary disease; respiratory inflammation;
KW	hypercholesterolaemia; crop protection agent; ds.

The invention relates to an isolated and purified nucleic acid segment comprising a desamine biosynthetic gene cluster, a fragment or its biologically active variant, where the nucleic acid sequence is not derived from the *eryC* gene cluster of *Saccharopolyspora erythraea* or *Streptomyces antibioticus*. The invention also relates to a macrolide biosynthetic gene cluster, or fragments thereof. The macrolide

biosynthetic gene cluster encodes proteins which synthesise methymycin, pikromycin, neomethymycin, narbomycin or a combination of these compounds. Recombinant or augmented cells comprising the desosamine and/or macrolide biosynthetic gene clusters are useful for the production of biologically active macrolides. The macrolide biosynthetic proteins are useful for synthesis of methymycin, pikromycin, neomethymycin and narbomycin. The alternative termination of polyketide synthesis may be useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA) monomers. The compounds produced by the recombinant host cells are useful as biopolymers, e.g., in packaging or biomedical applications, to engineer PHA monomer synthases or to prepare biologically active agents, such as chemotherapeutics, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macrolide-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polyketides in plants. The present sequence represents the macrolide biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC 15439. (Updated on 15-SEP-2003 to standardise OS field)

Seq	Sequence	37948 BP;	4901 A;	14718 C;	13311 G;	5018 T;	0 U;	0 Other;
Query Match	31.7%;	Score 158.6;	DB 3;	Length 37948;				
Best Local Similarity	59.0%;	Pred. No. 5.6e-17;						
Matches 291;	Conservative	0;	Mismatches 199;	Indels 3;	Gaps 1;			
QY	1	GGGGTCCCGACGCGGACGAGAACCGCATCTGTGTAAGCTGGTCCGCGGGCCACGCTTCG	60					
Db	32343	GGGCTCCCGACGCGGAGCGCGCGCGCGCTCTCACCTCTGTCGTACCAACGCGGCG	32402					
QY	61	ACGGTCTCGGCACACAGCGGCGCCGAAGGCATCGGCCCGCGCCAGCGGCTTCCAGAGAGTC	120					
Db	32403	GCGGTACTCGGCATCTCTCCGCCGACCGGGTGGCCCCCGGCGGTGCTTACCGAGCTC	32462					
QY	121	GGCTTCGACTCGCTGGCGCGGTCAACTTCGCGCAACAGCTCTGCACGCGGCCACCGGGCTG	180					
Db	32463	GGCTTCGACTCGCTGACGCGCGTGCAGCTCCGCAACCAAGCTCTCCACGGTGTGGCAAC	32522					
QY	181	CGGCTGCCGCGACGCTGATCTTCGACTACCCGACCCCGAGAGGCGCTGGTCCGCTACCTG	240					
Db	32523	AGGCTCCCGCGCACCAACGCGTCTTCACACACCGACGCCCGCGCAGCTCGCCGCGCACCTC	32582					
QY	241	CGGCTCGAACTCCTCGGGAGGCGCCGACACGAGCGCTTGAGCGGCGGGAAGACGACTCCGG	300					
Db	32583	CACGAGGCGTACTCTCGCACCGGCGGAGCGGCGCCCGACGAGCTGGAGGGGCGGGTGGCG	32642					
QY	301	CGAGTCTCTCGGCGCGTGGCTGTTCGCCCGGTTCAAGGAGCGGGCGTGTCTGGACACGCTG	360					
Db	32643	CGGGCGCTTGGCGCAACTGCCCTCGACCGGCTGGGGAGCGCGGGGTCTCTGACACCGTC	32702					
QY	361	CTCGGCTTCGCGACACCGGCACCGCAACCGGGCACGGACG---CCGAGACACCGAAGCG	417					
Db	32703	CTGCGGCTTACCGGATTCAGCGCCGAGCCGGTTCGGCGGTTCGGAACGGCGCGCGCC	32762					
QY	418	GCCCGGCGCGCGACGACGCGAGAACTGATCGACGCACTGGGACATCTCCGGTCTCGTGCAA	477					
Db	32763	GACCTTGGTCGGAGCCGGAGCGGTTCGATCGAGCACTTGGAGCCGAGGCGCTGATCCGG	32822					
QY	478	CGAGCCCTCGGGC	490					
Db	32823	ATGGCTCTCGGC	32835					

RESULT 5	
AAA75633	
ID AAA75633	standard; DNA; 38506 bp.
XX	
XX	
AC	AAA75633;
XX	
XX	
DT	22-JAN-2001 (first entry)
XX	
XX	





27-MAY-1999; 99US-00320878.  
(KOSA-) KOSAN BIOSCIENCES INC.  
Ashley G, Belach MC, Belach M, Medaniel R, Tang L;  
WPI; 2003-352291/33.  
Novel recombinant DNA compounds comprising coding sequences for  
PT desosamine transferase gene of Streptomyces venezuelae, useful for  
PT producing desosamine transferase which transfers desosamine to substrate  
XX polyketides.  
XX  
XX Claim 2; Col 21-54; 132pp; English.  
XX  
XX The present invention relates to recombinant DNA compounds that encode  
CC Streptomyces venezuelae (Sv) narbonolide polyketide synthases (PKSs). The  
CC recombinant PKSs are derived from narbonolide PKS and other genes  
CC involved in narbomycin and picromycin biosynthesis in recombinant host  
CC cells. The invention also discloses the S. venezuelae PKS gene cluster  
CC that results in the production of picromycin. Also disclosed are enzymes  
CC such as those responsible for glycosylation and hydroxylation, (e.g. C12  
CC hydroxylase (Pick)), desosamine biosynthesis, and desosaminyl transferase  
CC enzymes. The recombinant narbonolide, narbonolide derivatives, and  
CC polyketides are useful as antibiotics and as intermediates in the  
CC synthesis of compounds for pharmaceutical applications. The present  
CC sequence represents cosmid pKOS023-27 containing S. venezuelae PKS gene  
XX cluster.  
XX  
SQ Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 U; 0 Other;  
SQ  
Query Match 31.7%; Score 158.6; DB 7; Length 38506;  
Best Local Similarity 59.0%; Pred. No. 5.6e-17;  
Matches 251; Conservative 0; Mismatches 199; Indels 3; Gaps 1;  
QY 1 GCGGTCCCGACGCGGAGCAGAACCGCATCTCTGTGAAGCTGGTCCGCGGCCACGCTTCG 60  
Db 29315 GCGTCCCGGAGCCGAGCGCGCGCGTCTCTACCTCTGTCTACCGCGCGGCG 29374  
QY 61 AGGTGTCTGGCCACAGCGGGCCGGAAGGCAATCGGCCCGCGCCAGGCGTTCCAGAGATC 120  
Db 29375 GCGCTACTCGGCATCTCTCCCGGACCGGGTGGCCCGCGCGTCTTCAACGAGTCTC 29434  
QY 121 GAGTTCGACTCGCTGGCGCGGTCAACTCTCGGACACGCTGCACGGCGCCACCGGCTG 180  
Db 29435 GGCCTTCGACTCGCTGACGCGCGTGCAGTCTCCGCAACCGAGCTTCCACGGTGGTCCGCAAC 29494  
QY 181 CGGCTCCCGCGACGCTGATCTTCGACTACCCACCCCGGAGGGCGTGGTGGGTACCTG 240  
Db 29495 AGGCTCCCGCGACACACAGGTCTTGACACACCGAGCCCGCGCACTCGCGCGCACCTC 29554  
QY 241 CGGCTCGAATCTCTCGGGAGCCGACGACGCGCTTGAAGCGGCGGGAAGACGACTTCGG 300  
Db 29555 CACGAGGGGTACTCTCGCACCGGCGGAGCGCGCCCGACGAGCTGGAGGGGCGGGTGC 29614  
QY 301 CGAGTCTCTCGGGCGGTTCGCGCGGTTCAGAGAGGGCGGGGTGCTCGACACGCTG 360  
Db 29615 CGGGGCCCTGGCGAACTCGCCCTCGACCGGTGCGGGAACCGGGGGTCTCTCGACACCG 29674  
QY 361 CTCGCGCTCGCGACACCGGCAACCGGACCGGGCAGCGAGC--CCGAGACACCGAAGCG 417  
Db 29675 CTGCGCTCTCACGGGATCGAGCCCGAGCCCGGGTTCGGCGGTTCGAGCGGGGCGCCGCC 29734  
QY 418 GCCCGGGCGCGACGACGACGAACTGATCGACGACTGGGACATCTCCGGTCTCGTGCAA 477  
Db 29735 GACCTGTGTGGAGCGGAGCGGTGATCGACGACCTGGAACCGCGAGGCCCTGATCCGG 29794  
QY 478 CGAGCCCTCGGGC 490  
Db 29795 ATGGCTCTCGGCC 29807

ID	Accession	Name	Length	Score	DB	Mismatches	Indels	Gaps
AD	AAD55818	standard; DNA; 4725 BP.	4725	31.1%	DB 7;	0	3	1;
AC	AAD55818;							
AT	27-OCT-2003 (revised)							
DT	07-AUG-2003 (first entry)							
XX								
DE	Micromonospora carbonacea polyketide synthase (PKS) type I gene #8.							
KW	Macrolide; rosamycin; polyketide; polyketide synthase; PKS; enzyme;							
KM	gene; ds.							
OS	Micromonospora carbonacea.							
XX								
FH	Key	Location/Qualifiers						
CDS	1..4725	/tag= a						
FT	/product= "Polyketide synthase"							
XX								
PN	CA2391131-Al.							
XX								
PD	19-NOV-2002.							
XX								
PF	26-JUL-2002; 2002CA-02391131.							
XX								
PR	26-JUL-2001; 2001US-0307629P.							
XX	(ECOP-) ECOPIA BIOSCIENCES INC.							
PA								
PI	Yang X, Staffa A, Farnet CM;							
XX								
DR	WPI; 2003-343556/33.							
DR	P-PSDB; AAEE37002.							
XX								
PT	Novel isolated polypeptide involved in biosynthesis of macrolides by microorganisms, useful for biosynthesis of macrolides by microorganisms, preferably for biosynthesis of rosamycin.							
PT								
XX								
PS	Claim 1; Page 164-166; 206pp; English.							
CC	The invention relates to genes and proteins involved in the biosynthesis of macrolides by microorganisms. In particular it relates to the nucleic acids forming the biosynthetic locus for rosamycin (a 16-member macrolide antibiotic) from Micromonospora carbonacea. The invention is useful for the biosynthesis of macrolides by microorganisms. It allows direct manipulation of macrolides and related chemical structures by chemical engineering of the proteins involved in the biosynthesis of rosamycin. It is useful to catalyse certain biochemical reactions, in vitro or in vivo, to direct or enhance the synthesis or modification of a polyketide, polyketide substrate or its precursor. The present sequence is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-OCT-2003 to standardise OS field)							
SQ	Sequence 4725 BP; 482 A; 1782 C; 1885 G; 576 T; 0 U; 0 Other;							
Query Match	31.1%; Score 155.8; DB 7; Length 4725;							
Best Local Similarity	62.2%; Pred. No. 2.1e-16;							
Matches	263; Conservative							
QY	12 CGCGGAGCAGACCGCATCTCTGTGTCGGTTCGGCGGCACCGTTTCGACGGTGTCGG 71							
Db	4254 CGACGAACGCCGCGGGCTCTCTGACTGTTCGGCGGCAGGTTCGGCGGCTCTCGG 4313							
QY	72 CCACAGCGCGCGGAGGATCGCGCCCGCGCGGCGTTCAGAGGTTCGGCTTCGACTC 131							
Db	4314 CCACCGGGGCGCGAGCACGTTCGGCCCCGACGCCCGGTTCGGAGATTCGACTC 4373							
QY	132 GCTGGCGCGGTCACCTTCGCAACAGCCTGACCGCGGCACCGGGTTCGGCTTCGCGC 191							
Db	4374 GCTGACCGCGTCGACTTCGGCCAGCGGCTCAGGCGCGGGTTCGGCTTCGCTCGC 4433							
QY	192 GAGCTGATCTTCGACTACCCACCACCGCGAGCGCTGCTCGGCTACCTTCGCGTCACT 251							

FT CDS 47794..49083  
/\*tag= k  
/product= "Polyketide synthase #11"  
FT CDS 49092..49814  
/\*tag= l  
/product= "Polyketide synthase #12"  
FT CDS 49868..51226  
/\*tag= m  
/product= "Polyketide synthase #13"  
/transl\_except= (pos:51221..51223, aa:Xaa)  
/note= "Xaa corresponds to amino acids from position 452-1811"  
FT CDS 51506..53416  
/\*tag= n  
/product= "Polyketide synthase #14"  
/note= "CDS does not include start codon"  
FT CDS  
/partial  
complement(54569..53358)  
FT CDS  
/product= "Polyketide synthase #15"  
54997..56342  
/\*tag= p  
/product= "Polyketide synthase #16"  
56408..57634  
/\*tag= q  
/product= "Polyketide synthase #17"  
57657..59123  
/\*tag= r  
/product= "Polyketide synthase #18"  
/note= "CDS does not include start codon"  
FT CDS 59363..60196  
/\*tag= s  
/product= "Polyketide synthase #19"  
CA2391131-A1.  
XX  
19-NOV-2002.  
XX  
26-JUL-2002; 2002CA-02391131.  
XX  
26-JUL-2001; 2001US-0307629P.  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX Yang X, Staffa A, Farnet CM;  
XX  
XX WPI; 2003-343556/33.  
XX P-PSDB; AAE36995, AAE36996, AAE36997, AAE36998, AAE36999, AAE37000,  
XX AAE37001, AAE37002, AAE37003, AAE37004, AAE37005, AAE37006, AAE37007,  
XX AAE37008, AAE37009, AAE37010, AAE37011, AAE37012, AAE37013.  
XX  
XX Novel isolated polypeptide involved in biosynthesis of macrolides by  
XX microorganisms, useful for biosynthesis of macrolides by microorganisms,  
XX preferably for biosynthesis of rosamycin.  
XX  
XX Example 2; Page 59-94; 206pp; English.  
XX  
XX The invention relates to genes and proteins involved in the biosynthesis  
XX of macrolides by microorganisms. In particular it relates to the nucleic  
XX acids forming the biosynthetic locus for rosamycin (a 16-member  
XX macrolide antibiotic) from Micromonospora carbonacea. The invention is  
XX useful for the biosynthesis of macrolides by microorganisms. It allows  
XX direct manipulation of macrolides and related chemical structures by  
XX chemical engineering of the proteins involved in the biosynthesis of  
XX rosamycin. It is useful to catalyse certain biochemical reactions, in  
XX vitro or in vivo, to direct or enhance the synthesis or modification of a  
XX polyketide, polyketide substrate or its precursor. The present sequence  
XX is M. carbonacea polyketide synthase (PKS) type I gene cluster. (Updated  
XX on 27-OCT-2003 to standardise OS field)  
XX  
XX Sequence 60196 BP; 6934 A; 20042 C; 23822 G; 9398 T; 0 U; 0 Other;  
SQ

Query Match 31.1%; Score 155.8; DB 7; Length 60196;  
Best Local Similarity 62.2%; Pred. No. 1.4e-16;  
Matches 263; Conservative 0; Mismatches 157; Indels 3; Gaps 1;  
QY 12 GCGGAGCAAGACCGCATCTCTGCTGAAGCTGTCTCGCGGGCCACGCTTCGACGGTGTCTCGG 71  
Db 40545 CGACGAACGCGCGCGGGGCTGTCTCGACCTGTGTACGCGCGGAGGTTCGCGCGCTCTCGG 40604  
QY 72 CCACAGCGGCGCGCAAGAGCATCGGCGCGCGCGCTTCAGAGAGGTTCGACTTCGACTC 131  
Db 40605 CCACCGGGGCGCGGACGACGTTCGCGCGCGCGCTTCGCGGAGATCGGATTCGACTC 40664  
QY 132 GCTGGCGCGGTCAACCTCGCGCAACAGCTTCGACCGCGGCCACCGGGGTTCGCGGTTCGCCGC 191  
Db 40665 GCTGACCGCGGTCTGACCTTGGCCAAAGCGCTCAGGCGCGCGGTTCGCGGTTCGCCGC 40724  
QY 192 GAGCTGATCTTCGACTACCCCGGAGGCGCTGGTGGGTCTACCTGCGGTTCGACT 251  
Db 40725 CACCTCGTCTTCGACCAACCCCGGACCGCGGTTCGCGGTTCGCCGC 40784  
QY 252 CTTGCGGAGGCGCGACGACGCGCTTCGACGCGCGGCGGAGACGACCTTCGCGCGAGTCTCGC 311  
Db 40785 CGGTCCCG 40841  
QY 312 GCGCGTCCGTTCGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCG 371  
Db 40842 CGACCTGCGGTTCGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCG 40901  
QY 372 CGACACCGGCGCGACG 431  
Db 40902 GGGGCTGCG 431  
QY 432 CGA 434  
Db 40962 CGA 40964  
RESULT 11  
AAI99682 22/c  
Continuation (23 of 45) of AAI99682 from base 2200001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682  
WP Fragment Name Begin End  
WP AAI99682\_00 1 110000  
WP AAI99682\_01 100001 210000  
WP AAI99682\_02 200001 310000  
WP AAI99682\_03 300001 410000  
WP AAI99682\_04 400001 510000  
WP AAI99682\_05 500001 610000  
WP AAI99682\_06 600001 710000  
WP AAI99682\_07 700001 810000  
WP AAI99682\_08 800001 910000  
WP AAI99682\_09 900001 1010000  
WP AAI99682\_10 1000001 1110000  
WP AAI99682\_11 1100001 1210000  
WP AAI99682\_12 1200001 1310000  
WP AAI99682\_13 1300001 1410000  
WP AAI99682\_14 1400001 1510000  
WP AAI99682\_15 1500001 1610000  
WP AAI99682\_16 1600001 1710000  
WP AAI99682\_17 1700001 1810000  
WP AAI99682\_18 1800001 1910000  
WP AAI99682\_19 1900001 2010000  
WP AAI99682\_20 2000001 2110000  
WP AAI99682\_21 2100001 2210000  
WP AAI99682\_22 2200001 2310000  
WP AAI99682\_23 2300001 2410000  
WP AAI99682\_24 2400001 2510000  
WP AAI99682\_25 2500001 2610000  
WP AAI99682\_26 2600001 2710000  
WP AAI99682\_27 2700001 2810000  
WP AAI99682\_28 2800001 2910000  
WP AAI99682\_29 2900001 3010000  
WP AAI99682\_30 3000001 3110000



```
WP AAI19682_31 3210000
WP AAI19682_32 3200001
WP AAI19682_33 3310000
WP AAI19682_34 3410000
WP AAI19682_35 3500001
WP AAI19682_36 3600001
WP AAI19682_37 3700001
WP AAI19682_38 3800001
WP AAI19682_39 3900001
WP AAI19682_40 4000001
WP AAI19682_41 4100001
WP AAI19682_42 4200001
WP AAI19682_43 4300001
WP AAI19682_44 4400001

Query Match 30.8%; Score 154.4; DB 4; Length 110000;
Best Local Similarity 65.4%; Pred. No. 2.2e-16;
Matches 246; Conservative 0; Mismatches 121; Indels 9; Gaps 1;

QY 5 TCCCCGACCGCCGAGCAACCGCATCTCTGAAGCTGTCTCGCGCCACGCTTCGACGG 64
Db 95022 TGCCCGAAGACGAGCAACGCGCTCTCTGTGACCTGTGTGCTCGACATCGCCACCG 94963
QY 65 TGCTCGGCACAGCGCGCGCGAGGCATCGGCCCGCGCGCGGTTCGAGGAGTTCGGCT 124
Db 94962 TGCTGGGTAGCGCAGCCCGAAGCCATCGACCCGGATAGGGCATTCACAGGAGTTGGGT 94903
QY 125 TCGACTCGCTGCGCGCGGTCAACCTCGCAACAGCCTGCAAGCGGCGCGCGGTTCGGCG 184
Db 94902 TCGACTCGCTGCGCGCGGTCAACCTCGCAACAGCCTGCAAGCGGCGCGGTTCGGCG 94843
QY 185 TGCTCGGCACAGCGCGCGCGAGGCATCGGCCCGCGCGCGGTTCGAGGAGTTCGGCT 244
Db 94842 TTTCACCCAGCTCATCTTCGACTACCCCAATCCCGCGCGGTTCGCGGTATATGCGTC 94783
QY 245 TCGAATCTCT-----GCGGAGCGCGAGCGAGCGCTTCGACGCGCGGGAAGACACC 295
Db 94782 GAGAACTGCTCGGCTCATCAGCGCAAGACATTCAGCCGTGGCGCGCGGGAAGCGAAC 94723
QY 296 TCGGAGGAGTCTCGCGCGCGGTTCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCG 355
Db 94722 TACAAGCGCATTCGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCG 94663
QY 356 CGCTGCTCGGCTCGC 371
Db 94662 TGTTCGCTCGGCTGGC 94647

RESULT 12
AAI19683_22/c
Continuation (23 of 44) of AAI19683 from base 2200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI19683 Accession AAI19683
WP Fragment Name Begin End
WP AAI19683_00 1 110000
WP AAI19683_01 100001 210000
WP AAI19683_02 200001 310000
WP AAI19683_03 300001 410000
WP AAI19683_04 400001 510000
WP AAI19683_05 500001 610000
WP AAI19683_06 600001 710000
WP AAI19683_07 700001 810000
WP AAI19683_08 800001 910000
WP AAI19683_09 900001 1010000
WP AAI19683_10 100001 1110000
WP AAI19683_11 1100001 1210000
WP AAI19683_12 1200001 1310000
WP AAI19683_13 1300001 1410000
WP AAI19683_14 1400001 1510000
WP AAI19683_15 1500001 1610000
WP AAI19683_16 1600001 1710000
WP AAI19683_17 1700001 1810000
WP AAI19683_18 1800001 1910000
WP AAI19683_19 1900001 2010000
```

```
WP AAI19683_20 2100001
WP AAI19683_21 2210000
WP AAI19683_22 2300001
WP AAI19683_23 2400001
WP AAI19683_24 2500001
WP AAI19683_25 2600001
WP AAI19683_26 2700001
WP AAI19683_27 2800001
WP AAI19683_28 2900001
WP AAI19683_29 3000001
WP AAI19683_30 3100001
WP AAI19683_31 3200001
WP AAI19683_32 3300001
WP AAI19683_33 3400001
WP AAI19683_34 3500001
WP AAI19683_35 3600001
WP AAI19683_36 3700001
WP AAI19683_37 3800001
WP AAI19683_38 3900001
WP AAI19683_39 4000001
WP AAI19683_40 4100001
WP AAI19683_41 4200001
WP AAI19683_42 4300001
WP AAI19683_43 4400001

Query Match 30.8%; Score 154.4; DB 4; Length 110000;
Best Local Similarity 65.4%; Pred. No. 2.2e-16;
Matches 246; Conservative 0; Mismatches 121; Indels 9; Gaps 1;

QY 5 TCCCCGACCGCCGAGCAACCGCATCTCTGAAGCTGTCTCGCGCCACGCTTCGACGG 64
Db 97322 TGCCCGAAGACGAGCAACGCGCTCTCTGTGACCTGTGTGCTCGACATCGCCACCG 97263
QY 65 TGCTCGGCACAGCGCGCGCGAGGCATCGGCCCGCGCGGTTCGAGGAGTTCGGCT 124
Db 97262 TGCTGGGTAGCGCAGCCCGAAGCCATCGACCCGGATAGGGCATTCACAGGAGTTGGGT 97203
QY 125 TCGACTCGCTGCGCGCGGTCAACCTCGCAACAGCCTGCAAGCGGCGCGCGGTTCGCGCG 184
Db 97202 TCGACTCGCTGCGCGCGGTCAACCTCGCAACAGCCTGCAAGCGGCGCGGTTCGCGCG 97143
QY 185 TCGCGCGAGCGTGAATTCGACTACCCCAACCGCGAGCGCTGTGTGCTGCTGCTGCTG 244
Db 97142 TTTCACCCAGCTCATCTTCGACTACCCCAACTCGCGCGCGCTGTGCGCGCTATATG 97083
QY 245 TCGAATCTCT-----GCGGAGCGCGAGCGCGCTTCGACGCGCGGGAAGACACC 295
Db 97082 GAGAACTGCTCGGCTCATCAGCGCAAGACATTCAGCCGTGGCGCGCGGGAAGCGAAC 97023
QY 296 TCGCGCGAGTCTCGCGCGCGGTTCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCG 355
Db 97022 TACAAGCGCATTCGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCG 96963
QY 356 CGCTGCTCGGCTCGC 371
Db 96962 TGTTCGCTCGGCTGGC 96947

RESULT 13
AAI17184
ID AAI17184 standard; DNA; 65140 BP.
XX AAI17184;
XX AC
XX AAI17184;
XX DT 29-NOV-2001 (first entry)
XX DE Streptomyces noursei nysl DNA of nystatin PKS gene cluster.
XX DE Streptomyces noursei nysl DNA of nystatin; nystatin; PKS gene cluster;
XX KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX KW antifungal; antibiotic; nysl; ds.
XX OS Streptomyces noursei.
XX XX
```

PH	Key	Location/Qualifiers	DR	AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
FT	CDS	Complement(1..1035)	XX	
FT		/*tag= a	XX	New nystatin polyketide synthase polynucleotides and polypeptides, useful
FT		/product= "NysD2 partial protein"	FT	as antibiotics and antifungals.
FT	CDS	/note= "CDS does not include stop codon"	XX	
FT		complement(1056..2576)	PS	Claim 2; Page 116-151; 266pp; English.
FT		/*tag= b	XX	
FT		/product= "NysD1 protein"	CC	The present invention relates to the cloning and sequencing of the gene
FT	CDS	2806..6906	CC	cluster encoding a modular type I polyketide synthase (PKS) enzyme
FT		/*tag= c	CC	involved in the biosynthesis of the macrolide antibiotic nystatin. The
FT		/product= "NysA protein"	CC	nystatin PKS is useful as antifungal antibiotics. The present sequence is
FT	CDS	6952..16530	CC	a Streptomyces noursei nysI DNA of nystatin PKS gene cluster
FT		/*tag= d	XX	
FT		/product= "NysB protein"	SQ	Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 U; 0 Other;
FT	CDS	16550..49840		
FT		/*tag= e		Query Match 30.5%; Score 152.6; DB 4; Length 65140;
FT		/product= "NysC protein"		Best Local Similarity 60.7%; Pred. No. 4.6e-16;
FT	CDS	50260..51015		Matches 276; Conservative 0; Mismatches 164; Indels 15; Gaps 1;
FT		/*tag= f		
FT		/product= "NysE protein"	QY	1 GCGGTCCCGAGCGCGAGCAGAACCGCATCTGCTGAAGCTGGTCCGCGGCCACGGCTTCG 60
FT	CDS	51405..54305	Db	16030 GAGCTGTCCGCGCGAGCGCGCGGCCCTGGTCGAGCGGTCCGCGCGAGGCGTCC 16089
FT		/*tag= g		
FT		/product= "NysR1 protein"	QY	61 ACGGTGCTCGGCACACAGCGCGCGCGGAGGATCGGCGCCCGCAGCGCTTCCAGGAGGTC 120
FT	CDS	54329..57190	Db	16090 GCGACCTCGGCACAGACACCCCGAGCGCATCCCGCGCGGTGCTTCCGCGAGCTC 16149
FT		/*tag= h		
FT		/product= "NysR2 protein"	QY	121 GGTCTGACTGCTGGCGCGGTCAACCTCGGAAACAGCCTGACCGCGGCACCGGGCTG 180
FT	CDS	57180..59963	Db	16150 GGTCTGACTGCTGACCGCGGTGAACTCGCAACCGGTGCGCACCGCCCTCGGCCTG 16209
FT		/*tag= i		
FT		/product= "NysR3 protein"	QY	181 CGGTGCGCGGACGCTGATCTTCGACTACCCACCCCGAGGCGCTGTCGCTACCTG 240
FT	CDS	60415..61047	Db	16210 CGGTGCGCGGCGCGCTGCTGTTTCGACCCACCCCGAGCGCCCTCGCGCGCCCTC 16269
FT		/*tag= j		
FT		/product= "NysR4 (short) protein"	QY	241 CGGTGCGGAACTCTCGCGGAGGCGCGACGA-----CGGCTTGAAGCGGCGG 285
FT	CDS	61736..62497	Db	16270 GCGCGCTGCTCTTCGGCACCGCCCGGAGGACCGCGGACCGCGCGCCCGACGACCCC 16329
FT		/note= "CDS does not include start codon"		
FT		/*tag= k	QY	286 GAAGACGACTCTCGCGGAGTCTTCGCGCGCGTGCCTTCGCGCGGTTCGCGCGGTTCAGGAGGCGGCG 345
FT	CDS	60415..61047	Db	16330 GACGCGCGCATCGCGAGGCGCTCGCACCGTCCCATCGACGCGTTCGCGCAAGGCGGCG 16389
FT		/product= "ORF2 protein"		
FT		/note= "CDS does not include start codon"	QY	346 GTGCTGACACGCTGCTCGGCTCGCGCACCGCGCACCGACCGGCGACCGGCGACCGGCGAG 405
FT	CDS	63765..64961	Db	16390 CTCTCGACATGTGCTGAAACTCGCGCGAGCGGAGACCGACCGACCGCCCGCCCGAG 16449
FT		/*tag= m		
FT		/product= "ORF1 protein"	QY	406 ACCACCGAAGCGCGCCCGCGCGCGCGCGCGACGACGACGA 440
PN	WO200159126-A2.		Db	16450 GCGGACGCCCCCTCGGAATCCCTCGACGACATGGA 16484
PD	16-AUG-2001.			
PF	08-FEB-2001; 2001WO-GB000509.			
XX				
XX	08-FEB-2000; 2000GB-00002840.			
PR	10-APR-2000; 2000GB-00008786.			
PR	14-APR-2000; 2000GB-00009387.			
XX				
XX	(UYN0-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.			
PA	(SNTF) SINTEF STIFTELSEN IND TEK FORSK.			
PA	(ALPH-) ALPHARMA AS.			
PA	(SINV-) SINVENT AS.			
PA	(DZIE/) DZIEGLEWSKA H.			
PA	(ZOTC/) ZOTCHEV S B.			
PA	(SEKU/) SEKUROVA O N.			
PA	(FJAE/) FJAEVVIK E.			
PA	(BRAU/) BRAUTASET T.			
PA	(STRO/) STROM A R.			
PA	(VALL/) VALLA S.			
XX				
XX	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;			
PI	Valla S, Ellingsen TE, Sletta H, Gulliksen O;			
XX				
XX	WPI; 2001-557614/62.			
DR	P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,			

## RESULT 14

AAD17186

ID AAD17186 standard; DNA; 125401 BP.

XX

AC AAD17186;

XX

DT 29-NOV-2001 (first entry)

XX

DE Streptomyces noursei nystatin PKS gene cluster DNA.

XX

XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

XX antifungal; antibiotic; ds.

XX

OS Streptomyces noursei.

XX

FH Key Location/Qualifiers

FT CDS 6337..34771

FT

FT /\*tag= a

FT /product= "NysI complete protein"

FT 34792..51099

FT /\*tag= b

```

FT CDS /product= "NysJ protein"
FT 51155..57355
FT /*tag= c
FT /product= "NysK protein"
FT 57503..58687
FT /*tag= d
FT /product= "NysL protein"
FT complement(58786..58980)
FT /*tag= e
FT /product= "NysM protein"
FT /note= "CDS does not include start codon"
FT complement(59045..60241)
FT /*tag= f
FT /product= "NysN protein"
FT /note= "CDS does not include start codon"
FT complement(60238..61296)
FT /*tag= g
FT /product= "NysD2 complete protein"
FT 120628..121308
FT /*tag= h
FT /product= "NysR4 (long) protein"
FT
FT WO200159126-A2.
FT
FT 16-AUG-2001.
FT
FT 08-FEB-2001; 2001WO-GB000509.
FT
FT 08-FEB-2000; 2000GB-00002840.
FT 10-APR-2000; 2000GB-00008786.
FT 14-APR-2000; 2000GB-00009387.
FT
FT (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
FT (SNF) SINTEF STIFTELSEN IND TEK FORSK.
FT (ALPH-) ALPHARMA AS.
FT (SINV-) SINVENT AS.
FT (DZIE-) DZIELEWSKA H.
FT (ZOTC-) ZOTCHEV S B.
FT (SEKU-) SEKUROVA O N.
FT (FJAE-) FJAEVRIK E.
FT (BRAU-) BRAUTASET T.
FT (STRO-) STROM A R.
FT (VALL-) VALLA S.
FT
FT Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
FT Valla S, Ellingsen TE, Sletta H, Gulliksen O;
FT
FT WPI; 2001-557614/62.
FT P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
FT AAE10149, AAE10150.
FT
FT New nystatin polyketide synthase polynucleotides and polypeptides, useful
FT as antibiotics and antifungals.
FT
FT Claim 1; Page 188-254; 266pp; English.
FT
FT The present invention relates to the cloning and sequencing of the gene
FT cluster encoding a modular type I polyketide synthase (PKS) enzyme
FT involved in the biosynthesis of the macrolide antibiotic nystatin. The
FT nystatin PKS is useful as antifungal antibiotics. The present sequence is
FT a Streptomyces noursei nystatin PKS gene cluster DNA
FT
FT Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 U; 0 Other;
FT
FT Query Match 30.5%; Score 152.6; DB 4; Length 125401;
FT Best Local Similarity 60.7%; Pred. No. 4.2e-16;
FT Matches 276; Conservative 0; Mismatches 164; Indels 15; Gaps 1;
FT
FT 1 GCGGTCCCGACCGCAGCAGACCGCATCTGCTGAGCTGTCGCGCGCACGCTTCG 60
FT 76291 GAGCTGTCCGCGCGCAGCGCGCGCCCTGGTCGAGCGCTCGCGCGGCGTCC 76350
FT
FT 61 ACGGTGCTCGGCACAGCGCGCGCGAGGATCGGCGCGCGCATCGGCGCGCGGAGGTC 120

```

```

Db 76351 GCGACCTCGGCCACGACACCCCGAGCGCATCCCGCGCGCTGCTTCGCGAGGTC 76410
Qy 121 GGCTTCGACTCGCTGGCGCGGTCAACCTCGGCAACAGCCTGCAACGCGGCGCGGCTG 180
Db 76411 GGCTTCGACTCGCTGACCGCGCTGAACTGCGCAACCGGCTGCGCACCGGCTCGGCTG 76470
Qy 181 CGGCTGCGCGGAGCGCTGATCTTCGACTACCCACCGCGAGGCGGCTGGTTCGGTACCTG 240
Db 76471 CGGCTGCGCGGCGCTGCTGCTTCGACCAACCGCGCGGCGCTCGCGCGCACCTC 76530
Qy 241 CGGCTGCAACTCTCGCGGAGGCGCGAGA-----CGGCTCGAGCGGCGG 285
Db 76531 GCGCGCTGCTCTTCGCGCACCGCGCGAGACGCGCGCAACCGCGCGCGCGCGACCC 76590
Qy 286 GAAGACGACCTCGCGGAGTCTCGCGGCGCTGCGCTTCGCGCGCTTCAAGAGCGCGGC 345
Db 76591 GACGCGCGCATCGCGAGGCGCTCGCACCGCTCCCGCATCGGACGCTCGGCAAGCGGCG 76650
Qy 346 GTGCTGACACGCTGCTCGGCTCGCGGACACCGCGCAACCGCGGCGCACGAGCGCGAG 405
Db 76651 CTCCTCGACATGCTGTAACCTCGCGAGCGAGCGGACCGAGCGCGCGCGCGCGAG 76710
Qy 406 ACCACCGAAGCGCGCGCGCGCGCGCGAGCGAGCGAG 440
Db 76711 GCCGACGCGCGCTCGGAATCGCTCGACGACATGGA 76745

RESULT 15
AAD55817
ID AAD55817 standard; DNA; 11238 BP.
XX
AC AAD55817;
XX
DT 27-OCT-2003 (revised)
DT 07-AUG-2003 (first entry)
XX
DE Micromonospora carbonacea polyketide synthase (PKS) type I gene #7.
XX
KW Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme;
KW gene; ds.
XX
OS Micromonospora carbonacea.
XX
FH Key Location/Qualifiers
FT 1..11238
FT /*tag= a
FT /product= "polyketide synthase"
FT /note= "CDS does not include start codon"
FT /partial
XX
CA2391131-Al.
XX
PD 19-NOV-2002.
XX
PF 26-JUL-2002; 2002CA-02391131.
XX
PR 26-JUL-2001; 2001US-0307629P.
XX
PA (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI Yang X, Staffa A, Farnet CM;
XX
DR WPI; 2003-343556/33.
DR P-PSDB; AAE37001.
XX
PT Novel isolated polypeptide involved in biosynthesis of macrolides by
PT microorganisms, useful for biosynthesis of macrolides by microorganisms,
PT preferably for biosynthesis of rosamycin.
XX
PS Claim 1; Page 151-158; 206pp; English.
XX
CC The invention relates to genes and proteins involved in the biosynthesis

```

CC of macrolides by microorganisms. In particular it relates to the nucleic  
CC acids forming the biosynthetic locus for roaramycin (a 16-member  
CC macrolide antibiotic) from Micromonospora carbonacea. The invention is  
CC useful for the biosynthesis of macrolides by microorganisms. It allows  
CC direct manipulation of macrolides and related chemical structures by  
CC chemical engineering of the proteins involved in the biosynthesis of  
CC roaramycin. It is useful to catalyze certain biochemical reactions, in  
CC vitro or in vivo, to direct or enhance the synthesis or modification of a  
CC polyketide, polyketide substrate or its precursor. The present sequence  
CC is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-  
CC OCT-2003 to standardise OS field)  
XX  
SQ Sequence 11238 BP; 1110 A; 3596 C; 4707 G; 1825 T; 0 U; 0 Other;  
Query Match 29.7%; Score 148.8; DB 7; Length 11238;  
Best Local Similarity 58.6%; Pred. No. 2.4e-15;  
Matches 258; Conservative 0; Mismatches 182; Indels 0; Gaps 0;  
QY 1 GCGGTCCCGAGCGCGAGCAGAACCGCATCTCTGAAGCTGTCCGGCGCCACCGCTTCG 60  
Db 10717 GGGCTGACCGTGGCCGAAACAGCGGGAGCTGGTCTGGAGGCGTGGCGGACACCGGGCG 10776  
QY 61 AGGFTGCTCGGCACAGCGCGCCGAGGATCGGCCCGCGCGGCTTCCAGAGGTC 120  
Db 10777 GCGGTCTCGGACACGCCACCGCGAGCGCTCGACCGCGGCGCTTCCGGGAACTC 10836  
QY 121 GGCCTTCGACTCGTGGCGCGGTCACCTCCGCAACAGCGCTGACCGCGGCCACCGGGCTG 180  
Db 10837 GGCCTTCGACTCGTGGCGCGGTCGCAATCGGCTGGCCACCGCGTCCGGGCTG 10896  
QY 181 CGGCTGCCCGGAGCGCTGATCTTCGACTACCCACCGCGGAGGCGCTGGTCCGGTACCTG 240  
Db 10897 CGCCTGCCCGGAGCGCTGCTTCGACCAACCCCGGAGCGCTGGCGGAGCACCTG 10956  
QY 241 CGGCTCGAATCTCGCGGAGCGCGACCGGCTGGACGGCGGGAAGACGACCTCCGG 300  
Db 10957 CTCGCCGGGCTCGCGCCCGAGAGCGGCCCGGGCCGAGTTGCCGTTGTCGCCGAGCTGGGC 11016  
QY 301 CGAGTCTCTCGCGGCCGCTGCGCTTCGCCCGGTTCAAGGAGGCGGGCGGTGTGGACACGCTG 360  
Db 11017 CGGCTGGAGGCGGCCCTGGCGGCCACCGACGGGGCGGCCCTCGACGGGCTGGACGACCTG 11076  
QY 361 CTCGGGCTTCGCGACACCGGACCGGAACCGGGGACGAGACCGGAGACCGGAGCGGCC 420  
Db 11077 GTCCGCGGGAGGTGGGCGTCCGATCGCGCGCTGGCCCGCAGGTGGGGCGGGCCGGC 11136  
QY 421 CCGGCGCGCGACGACGCGA 440  
Db 11137 GACGACGTGGCGGCGACGGA 11156

```

OM nucleic - nucleic search, using sw model

Run on:      June 16, 2004, 23:22:28 ; Search time 54.6667 Seconds
              (without alignments)
              5085.922 Million cell updates/sec

Title:       US-10-042-665A-3_COPY_15043_15543
Perfect score: 501
Sequence:    1 gcggtccgcagccgagca.....ccctcgggcagacgactga 501

Scoring table:  IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

```

Total number of bits satisfying chosen parameters: 1365418

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```
Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	158.6	31.7	4689	3	US-09-105-537-34	Sequence 34, Appl
2	158.6	31.7	36778	3	US-09-105-537-5	Sequence 5, Appli
3	158.6	31.7	38506	3	US-09-320-878-19	Sequence 19, Appl
4	158.6	31.7	38506	4	US-09-141-908-1	Sequence 1, Appli
5	158.6	31.7	38506	4	US-09-657-440-19	Sequence 19, Appl
C 6	154.4	30.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 7	154.4	30.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
8	140.2	28.0	11220	3	US-09-105-537-32	Sequence 32, Appl
9	135.8	27.1	43280	2	US-08-804-227C-1	Sequence 1, Appli
10	132	26.3	44377	2	US-08-804-227C-7	Sequence 7, Appli
11	132	26.3	44377	2	US-08-804-198-1	Sequence 1, Appli
12	130.8	26.1	11219	1	US-07-642-734C-1	Sequence 1, Appli
13	130.8	26.1	11219	3	US-08-439-009A-1	Sequence 1, Appli
14	127.4	25.4	50937	3	US-09-428-517-1	Sequence 1, Appli
C 15	126.2	25.2	561	3	US-09-154-083-13	Sequence 13, Appl
16	125.6	25.1	47981	4	US-09-679-279-1	Sequence 1, Appli
17	124.8	24.9	1681	4	US-09-434-288-7	Sequence 7, Appli
18	124.4	24.8	77536	4	US-09-410-551B-1	Sequence 1, Appli
C 19	122	24.4	77536	4	US-09-410-551B-1	Sequence 1, Appli
20	118.4	23.6	4466	4	US-09-410-551B-20	Sequence 20, Appl
21	118.4	23.6	4478	4	US-09-410-551B-16	Sequence 16, Appl
22	118.4	23.6	4547	4	US-09-410-551B-22	Sequence 22, Appl
23	118.4	23.6	4571	4	US-09-410-551B-18	Sequence 18, Appl
24	117.2	23.4	80161	3	US-09-036-987A-1	Sequence 1, Appli
25	117.2	23.4	80161	3	US-09-370-700-1	Sequence 1, Appli
26	117.2	23.4	80161	3	US-09-603-207-1	Sequence 1, Appli
27	117	23.4	20235	1	US-07-642-734C-3	Sequence 3, Appli

4543 CTGCGCCTCACCGGATCGAGCCGAGCCGGTTCGGCGGTTCCGAGCGGCGCGCC 4602  
418 GCCCGGCGCGCAGCAGCACTGATCGACGCACCTCGGTCCTCGTGCA 477  
4603 GACCTGGTTCGAGCGGAGCGTCGATCGACGACCTTGAGCGGAGGCCCTGATCGG 4662  
478 CGAGCCTTCGGC 490  
4663 ATGGGTCTCGGC 4675  
Db

```

RESULT 2
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

```

```

RESULT 3
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1998-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-320-878-19

```





GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 30.8%; Score 154.4; DB 3; Length 4403765;  
Best Local Similarity 65.4%; Pred. No. 6e-20;  
Matches 246; Conservative 0; Mismatches 121; Indels 9; Gaps 1;  
1;  
QY 5 TCCCGGAGCGGCGGAGCAACCGCATCTCTGCTGAAGCTGTTCGGCGCCACGCTTCGACGG 64  
Db 2297322 TGCCCGAAGACGAGCAACACACGCGCTCTCTGTGGACCTGTGTGGCTGCGCATATCGCCACCG 2297263  
QY 65 TGCTCGGCACAGCGCGCGGAGGAGCATCGGCCCGCGCCAGCGTTTCCAGGAGGTTCGGCT 124  
Db 2297262 TGCTGGGTAGCCCGAGCCCGGAGCCATCGACCCGGATAGGGCAATTCAGGAGTTGGGTT 2297203  
QY 125 TCGACTCGTGTGCGCGGCTCAACCTCCGCAACAGACCTCGACCGCGCCACCGGGCTTCGGCG 184  
Db 2297202 TCGACTCGCTACCGCGGTGGAATTCGCAACCGGCTCAATTCGCGCCACCGGCTTGGCGC 2297143  
QY 185 TCCCGGAGCGTGTGATCTTCGACTACCCACCGCGGAGCGCTTCAAGGAGCGCGCTTCGCGG 244  
Db 2297142 TTTCACCCACGCTCATCTTCGACTACCCCAACTCCCGCGCGCTGCGCGCTATATGCGTC 2297083  
QY 245 TCGAATCTCT-----GCGGGAGCGGAGCGGCTGACGCGGCGGAGGAGGACGACC 295  
Db 2297082 GAGAACTGCTCGGCTCATCCGCAAGACACTTCAGCCGTGGCGCGCGGAGGAGCGGAAC 2297023  
QY 296 TCCGGCGAGTCTCGCGCGCGTCCCGTTCGCGCGGTTCAAGGAGCGCGGCTTCGCTGGACA 355  
Db 2297022 TACAAGGCATTGTGGCGTCCATTCCCGGTCAAGCGCTTACGGCAGGCGGAGTGTGGATC 2296963  
QY 356 CGCTGCTCGGCTCGC 371  
Db 2296962 TGTTCGCTCGGCTGGC 2296947

RESULT 7  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529

TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 30.8%; Score 154.4; DB 3; Length 4411529;  
Best Local Similarity 65.4%; Pred. No. 6e-20;  
Matches 246; Conservative 0; Mismatches 121; Indels 9; Gaps 1;  
1;  
QY 5 TCCCGGAGCGGCGGAGCAACCGCATCTCTGCTGAAGCTGTTCGGCGCCACGCTTCGACGG 64  
Db 2295022 TGCCCGAAGACGAGCAACACACGCGCTCTCTGTGGACCTGTGTGGCTTCGACATCGCCACCG 2294963  
QY 65 TGCTCGGCACAGCGCGCGGAGGAGCATCGGCCCGCGCCAGCGTTTCAGGAGGTTCGGCT 124  
Db 2294962 TGCTGGGTAGCCCGAGCCCGGAGCCATCGACCCGGATAGGGCAATTCAGGAGTTGGGTT 2294903  
QY 125 TCGACTCGTGTGCGCGGCTCAACCTCCGCAACAGACCTTCGACCGCGCCACCGGGCTTCGCGC 184  
Db 2294902 TCGACTCGCTACCGCGGTGGAATTCGCAACCGGCTCAATTCGCGCCACCGGCTTGGCGC 2294843  
QY 185 TCCCGGAGCGTGTGATCTTCGACTACCCACCGCGGAGCGCTGTTCGGCTACCTTCGCGG 244  
Db 2294842 TTTCACCCACGCTCATCTTCGACTACCCCAACTCCCGCGCGCTGCGCGCTATATGCGTC 2294783  
QY 245 TCGAATCTCT-----GCGGGAGCGGAGCGGCTGACGCGGCGGAGGAGGACGACC 295  
Db 2294782 GAGAACTGCTCGGCTCATCACCAGCAAGACACTTCAGCCGTGGCGCGGAGGAGCGGAAC 2294723  
QY 296 TCCGGCGAGTCTTCGCGCGCGCTGCGCGTTTCGCGCGGTTCAAGGAGCGCGGCTTCGCTGGACA 355  
Db 2294722 TACAAGGCATTGTGGCGTCCATTCCGGTCAAGCGCTTACGGCAGCGGAGTGTGGATC 2294663  
QY 356 CGCTGCTCGGCTCGC 371  
Db 2294662 TGTTCGCTCGGCTGGC 2294647

RESULT 8  
US-09-105-537-32  
; Sequence 32, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 32  
; LENGTH: 11220  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-32

Query Match 28.0%; Score 140.2; DB 3; Length 11220;  
Best Local Similarity 64.1%; Pred. No. 2.6e-17;  
Matches 211; Conservative 0; Mismatches 118; Indels 0; Gaps 0;  
0;  
QY 1 GGGGTCCCGGAGCGGAGGAGCAACCGCATCTCTGCTGAAGCTGTTCGGCGCCACGCTTCG 60  
Db 4309 GGGCTCACCGCGCGCGGAGCAACAGCAGCGCGCTATATGAGGAGTTGGTTCGCGAGCACCTTCGCC 4368  
QY 61 ACGGTGCTCGGCCACAGCGCGCGGAGGAGCATCGGCCCGCGCGAGCGCTTCAGGAGGTC 120  
Db 4369 GTGGTCTCTCAACACACCCCTCCCGGAGCGCGTTCGACACCGGGCGGCGCTTCGCTGACCTC 4428  
QY 121 GGCTTCGACTCGCTGGCGCGGCTCAACCTCCGCAACAGCTTCGACGCGCCACCGGCTG 180

Db 4429 GGATTGACTCGTACGCGGCTGAGCTCCGACACCGCTCAAGAACGCCACCGGCTG 4488  
QY 181 CGGTGCGCGGACGCTGATCTTGACTACACCCACCGGAGGCGCTGGTGGCTACCTG 240  
Db 4489 GCGCTCCCGGCACTCTGCTCTTCTGACTACCGACCCCGGACGCTGGCGGAGTTCTCTC 4548  
QY 241 CGGTGCAACTCTCTGCGGAGCGGACGACGCTGAGCGGCGGAGACGACCTCCGG 300  
Db 4549 CTCGCGGAGATCTCTGGCGAGCAGCGCTGCGCGGAGAGCTTCCGCTGAGCGCGG 4608  
QY 301 CGAGTCTCTCGGCGGCTGCGCTTCCGCCG 329  
Db 4609 GTCGACGAGAGCCGCTCGGATGCTGG 4637

## RESULT 9

US-08-804-227C-1  
; Sequence 1, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43280 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 816..14234  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14351..19945  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20010..31199  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31232..36067  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36249..41774  
US-08-804-227C-1

Query Match

27.1%; Score 135.8; DB 2; Length 43280;

Best Local Similarity 57.0%; Pred. No. 1.5e-16;  
Matches 248; Conservative 0; Mismatches 187; Indels 0; Gaps 0;  
QY 16 GAGCAGAACCGATCTCTGTAAGCTGGTCCGCGCCACAGCTTCAGACGCTGCTCGGCCAC 75  
Db 35510 GAGCGCGAGGAGATCTGCTCGGCCCTGGTGGCCCGGCAATGTGGCCGCCGTACTCGGCCAC 35569  
QY 76 AGCGCGCGGAGGAGGATCGGCCCGCGCCAGGCGTTCCAGGAGGTTCGGCTTCGACTCGCTG 135  
Db 35570 CCGGGGACCGCGGACATCGTCCGACCGCTGTTTCNAGGAGCTGGGGTTCAGTTTCGCTC 35629  
QY 136 GCCGCGGTCAACCTCCGCAACAGCTGCAACGCGGCCACCGGGGCTGGGGTTCGCCGCGACG 195  
Db 35630 ACCGCGGTGAGCTGGCCCGGGCGGTGGGCCGCGGAGTGGGACGGAAGCTGCGCGCGACG 35689  
QY 196 CTGATCTTCGACTACCCACCCCGGAGGCGCTGGTGGCTTACCTCGCACTTCCTG 255  
Db 35690 CTGGTCTTCACCATCCGACTGCGCGCGCGCGCTGAAACACCTGGCGGAGTGTGTGACA 35749  
QY 256 CGGAGGCGGACGACGCGGCTTGACGCGGCGGGAAGACGACCTCGGCGAGTCTTCGCGGCC 315  
Db 35750 CGCGCGCGCGGTCCCGCGCGCTCCCGCGGAGGAGGAGCGCGCGCGCGCGCGCGCG 35809  
QY 316 GTGCCGTTCCCGCGGTTCAGAGAGGCGGCGGTGCTGGACACGCTGCTCGGCTTCGCGGAC 375  
Db 35810 GTGCCGCTCGAACCGGCTGAGGGGAAGCGGCGCTGTGAGCGCACTGCTGGGCTTCGCGG 35869  
QY 376 ACCGCGACCGAACCGGCGGACGCGGAGACGCGGAGACACCGAGCGGCCCGCGCGCGGCGG 435  
Db 35870 GACGAATCCGGGCGGACAAACCCCGCTACGCTGCGCGCGTCCGCGGACACCCCGGCGG 35929  
QY 436 GCAGAACTGATCGAC 450  
Db 35930 GAGGACCGGACGCG 35944

## RESULT 10

US-08-804-227C-7  
; Sequence 7, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44377 base pairs  
TYPE: nucleic acid









Db 10905 ACGTGGTCTTCGACACCCGAAAGCCTCCGGGTCCGGTTTCTCGACGCCGAGCTC 10964  
QY 253 -----CTCGGGAGGCGAGCAGCGGCTGACCGGGGGAGAC 291  
Db 10965 GGCACCGAGGTCCGGGGAGGCGCGTCCGCCCTCGCGGCTGACCGCTGGAAGGC 11024  
QY 292 GACTCCGGCGAGTCTCCGGCGCGTCCGGCTTCGCCCGGTTCAG--GAGGCGGCGGTG 348  
Db 11025 GCCTCCGCGAGTCCCGCAACCGAGCGGAGAGCTGTACAGCGCTTGGACCGATG 11084  
QY 349 CTGGACACGCTGCTCGGCTTCGCCGACACCGGACCGAACCGGCGACCGCGAGACC 408  
Db 11085 CTCGCGCGCTACGCGCGGTCCGCCGCGCGCGCGCTCCGGGACCGGCGCAACCCG 11144  
QY 409 ACCGAGCGCGCCCGCC--GCCGACGACGCGAAGTGTACACGCACTGGAC 459  
Db 11145 TCCGGCGACACTGGCGGAGGCGGCGTGGACGAACCTGCTCGAAGCACTCGGC 11198

## RESULT 14

US-09-428-517-1  
; Sequence 1, Application US/09428517  
; Patent No. 6251636  
; GENERAL INFORMATION:  
; APPLICANT: Shah, Sanjay Krishnakant  
; APPLICANT: McDaniel, Robert  
; APPLICANT: Tang, Li  
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 30062-20029.00  
; CURRENT APPLICATION NUMBER: US/09/428,517  
; CURRENT FILING DATE: 1999-10-28  
; EARLIER APPLICATION NUMBER: 60/120,254  
; EARLIER FILING DATE: 1999-02-16  
; EARLIER APPLICATION NUMBER: 60/106,100  
; EARLIER FILING DATE: 1998-10-29  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 50937  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA  
US-09-428-517-1

Query Match 25.4%; Score 127.4; DB 3; Length 50937;  
Best Local Similarity 62.3%; Pred. No. 4.8e-15;  
Matches 200; Conservative 0; Mismatches 121; Indels 0; Gaps 0;  
QY 5 TCCCGGACGCGGAGCAGAACCGATCTCTGAGCTGTCCGGCGGCGACGCTTCGACGG 64  
Db 13102 TCACCGCGGTCCGACGAGAACGTTCTTGTGTGAGCTCGTGAGGAGCAGCGCGCGCTCG 13161  
QY 65 TGCTCGGCGCACAGCGGCGCGGAGGATCGGCGCGCGCGGCTTCAGGAGGTCCGCT 124  
Db 13162 TCCTCGACATTCGGGCGCGGAGCGGTGCTCGGACCGGAGCTTCAAGGATTCGGAT 13221  
QY 125 TCGACTCGCTGGCGCGGTCAACTCCGCAACAGCTTCGACGCGGCGCACCGGCGTCCGGC 184  
Db 13222 TCGACTCGCTGACCTCGGTGAACTGGCAACCGGCTGAACACCGCCACCGGCTTCAGC 13281  
QY 185 TGCCCGGACGCTGATCTTCGACTACCCACCGGAGGCGCTGTTGGCTTACCTGCGCG 244  
Db 13282 TGCCCGTGACGCGCGCTTCGACTACGCGAGCGCGCGCTGCGCGCCATCTCGCT 13341  
QY 245 TCGAACTCTTCGGGAGGCGCGACGCGCTTCGAGCGCGGAGAGACGACCTCCGGCGAG 304  
Db 13342 CCAGGCTGATGACGACGATGGTGACACCGTGTCTTCCCGCGCTGAGAGACCGCGA 13401  
QY 305 TCCTTCGGGCGGTGCCGTTCCG 325  
Db 13402 TCGACGAGCGGATCGCGATCG 13422

RESULT 15  
US-09-154-083-13/c  
; Sequence 13, Application US/09154083  
; Patent No. 6150513  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Kai  
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA  
; TITLE OF INVENTION: Constructs Therefor  
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz  
; CURRENT APPLICATION NUMBER: US/09/154,083  
; CURRENT FILING DATE: 1998-09-16  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 561  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus  
US-09-154-083-13

Query Match 25.2%; Score 126.2; DB 3; Length 561;  
Best Local Similarity 67.0%; Pred. No. 9.4e-15;  
Matches 179; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
QY 16 GAGCAGAACCGCATCTCTGTAAGCTGTCCGGCGGCGACGCTTCGACGCTCGGCCAC 75  
Db 398 GAGCGGCGGCACACTCTGTGGAGCTGTTCGGGACGAGGCGGATCGGCTTCGACTCGCTG 339  
QY 76 AGCGGCGCGGAGGAGCATCGCCCGCGCGCTTCAGGAGTTCGAGGTCGCTTCGACTCGCTG 135  
Db 338 GGCTCCGAGGCGCGCTGTGACGCGCAAGGCGTTCGGGAGATCGGCTTCGACTCGCTG 279  
QY 136 GCGCGGTCAACTCTCGCAACAGCTTCGACGCGGCGGCGGCTTCGGCTTCGAGTTCGCTG 195  
Db 278 ACCGCGGTTCGAACTTCGCGAACCGGCTGGATACCGGCAACCGGCGCTTCGCGGCGAGC 219  
QY 196 CTGATCTTCGACTACCCACCGCGGCGCTGTTCGGCTACCTTCGCGCTCGAACTCTCTG 255  
Db 218 CTGCTGTTCGACTACCGGCGCGCGCTTCGGCGCGCTTCGGGCGGCTTCGGTTCGGAGCTCTC 159  
QY 256 CGGAGGCGCGGAGCAGCGGCTTCGACGGG 282  
Db 158 GCGGGGAGGCGGATGCGGTGCGCGCG 132

Search completed: June 17, 2004, 03:22:09  
Job time : 68.6667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:31:09 ; Search time 293.667 Seconds

(without alignments)  
7787.123 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543

Perfect score: 501

Sequence: 1 gcggtcccgagcgcgagca.....ccctcgggcagcagctga 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2998549 seqs, 2282253817 residues

Total number of hits satisfying chosen parameters: 5997098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	158.6	31.7	4689	9 US-09-861-289-34	Sequence 34, Appl
2	158.6	31.7	4689	9 US-09-860-846-34	Sequence 34, Appl
3	158.6	31.7	4689	10 US-09-988-384B-34	Sequence 34, Appl
4	158.6	31.7	4689	10 US-09-836-821-34	Sequence 34, Appl
5	158.6	31.7	4689	15 US-10-271-889-34	Sequence 34, Appl
6	158.6	31.7	36778	9 US-09-861-289-5	Sequence 5, Appl
7	158.6	31.7	36778	9 US-09-860-846-5	Sequence 5, Appl
8	158.6	31.7	36778	15 US-10-271-889-48	Sequence 48, Appl
9	158.6	31.7	37948	10 US-09-988-384B-5	Sequence 5, Appl
10	158.6	31.7	37948	10 US-09-988-384B-5	Sequence 5, Appl
11	158.6	31.7	38506	10 US-09-793-708-19	Sequence 19, Appl
12	158.6	31.7	38506	15 US-10-201-365-1	Sequence 1, Appl
13	158.6	31.7	38506	15 US-10-160-539-19	Sequence 19, Appl
14	155.8	31.1	4725	15 US-10-205-032-17	Sequence 17, Appl

15	155.8	31.1	60196	15 US-10-205-032-1	Sequence 1, Appl
16	153.6	30.7	5721	15 US-10-156-761-2880	Sequence 2880, Ap
17	153.6	30.7	125746	15 US-10-156-761-15102	Sequence 15102, A
18	153.6	30.7	9025608	15 US-10-156-761-1	Sequence 1, Appl
19	148.8	29.7	11238	15 US-10-205-032-15	Sequence 15, Appl
20	147.2	29.4	18438	15 US-10-156-761-2886	Sequence 2886, Ap
21	147.2	29.4	125746	15 US-10-156-761-15102	Sequence 15102, A
22	147.2	29.4	9025608	15 US-10-156-761-1	Sequence 1, Appl
23	141	28.1	18435	15 US-10-156-761-412	Sequence 412, App
24	141	28.1	100000	15 US-10-156-761-15103	Sequence 15103, A
25	140.8	28.1	14055	15 US-10-156-761-2883	Sequence 2883, Ap
26	140.2	28.0	11220	9 US-09-861-289-32	Sequence 32, Appl
27	140.2	28.0	11220	9 US-09-860-846-32	Sequence 32, Appl
28	140.2	28.0	11220	10 US-09-988-384B-32	Sequence 32, Appl
29	140.2	28.0	11220	10 US-09-836-821-32	Sequence 32, Appl
30	140.2	28.0	11220	15 US-10-271-889-32	Sequence 32, Appl
31	139.6	27.9	18717	15 US-10-156-761-927	Sequence 927, App
32	139.6	27.9	30690	17 US-10-204-862A-1	Sequence 1, Appl
33	136.6	27.3	84428	17 US-10-229-148B-1	Sequence 1, Appl
34	135.8	27.1	11817	15 US-10-156-761-2884	Sequence 2884, Ap
35	135.8	27.1	23238	15 US-10-156-761-415	Sequence 415, App
36	134.4	26.8	64492	13 US-10-378-083-1	Sequence 1, Appl
37	134.2	26.8	3072	13 US-10-282-122A-25666	Sequence 25666, A
38	131.4	26.2	85692	17 US-10-461-194-1	Sequence 1, Appl
39	131.2	26.2	10692	15 US-10-156-761-414	Sequence 414, App
40	130.8	26.1	86941	17 US-10-461-194-2	Sequence 2, Appl
41	130	25.9	10839	15 US-10-156-761-2882	Sequence 2882, Ap
42	130	25.9	11910	15 US-10-156-761-2879	Sequence 2879, Ap
43	128.4	25.6	5505	15 US-10-156-761-413	Sequence 413, App
44	127.4	25.4	50937	10 US-09-808-880-1	Sequence 1, Appl
45	126.4	25.2	13494	15 US-10-156-761-2355	Sequence 2355, Ap

# ALIGNMENTS

RESULT 1  
US-09-861-289-34  
; Sequence 34, Application US/09861289  
; Patent NO. US20020110897A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/861,289  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ IDS NOS: 43  
; SOFTWARE: FastSeq For Windows Version 3.0  
; SEQ ID NO 34  
; LENGTH: 4689  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-861-289-34

Query Match	31.7%	Score 158.6;	DB 9;	Length 4689;
Best Local Similarity	59.0%	Pred. No. 1.1e-27;		
Matches 291;	Conservative	0;	Mismatches 199;	Indels 3; Gaps 1;
QY	1	GGGTCCTCCGAGCGCGAGCAGAACCGCATCTGCTGAGCTGTCGGCGGCGCACGCTTCG	60	
Db	4183	GGGTCCTCCGAGCGCGAGCGCGCGGCGGCTCTCACCTCGTCGACCGCGCGG	4242	
QY	61	AGGTCCTCCGAGCGCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	120	
Db	4243	GGCCTACTCGGCGCATCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	4302	
QY	121	GGCTTCGACTCGCTGCGCGCGCGCTCAACTCTCCGCAACAGCGCTCGACGCGCGCGCGCG	180	

Db 4303 GGCTTGGACTCGCTGACGGCGGTGAGTCCGCAACCAAGCTCTCCACGGTGGTCCGCAAC 4362  
Qy 181 CGGTGCGCGGAGCGCTGACTTGGATACCCACCCCGGAGGGCGTGGTGGGTACTCTG 240  
Db 4363 AGGCTCCCGCGCACACCGGTCTTGGACACCCGACCGCGCGCGCACTGCGCGGCACTC 4422  
Qy 241 CGCGTCAAACTCTTCGCGGGAGCGGACGACGCGCTTGAACGGCGGGGGAAGACGACTCCGG 300  
Db 4423 CACGAGCGGTACTCTGCACCGCGCGAGCGCGCCCGGACGACTGGAGGGGGGGTGGCG 4482  
Qy 301 CGAGTCTCGCGCGGTGCGCGCTTCCCGCGGTTCAGGAGGGCGGCGTGGGACACGCTG 360  
Db 4483 CGGCGCTTGGCGGAATCGCCCTCGACCGGTGCGGGAGCGCGGGGTCTTCCGACACGCTC 4542  
Qy 361 CTGGCGCTCGCGCACACCGGACCGAACCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417  
Db 4543 CTGGCGCTCGCGCACACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4602  
Qy 478 CGAGCCCTCGGGC 490  
Db 4663 ATGGCTCTCGGCC 4675

## RESULT 2

US-09-860-846-34  
; Sequence 34, Application US/09860846  
; Patent No. US20020164742A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/860.846  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SEQ ID NO 34  
; LENGTH: 4689  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-860-846-34

Query Match 31.7%; Score 158.6; DB 9; Length 4689;  
Best Local Similarity 59.0%; Pred. No. 1.1e-27;  
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;  
Qy 1 GCGGTCCCGACGCGCGAGCAACCGCATCTCTGTAAGCTGGTCCGCGGCGCACTTCG 60  
Db 4183 GCGCTCCCGAGCGCGGCGCGCGGTCTCTCACTCCCTCGTCCGTACCCACGCGCG 4242  
Qy 61 ACGGTGCTCGGCCACAGCGGCGCGAGGCGATCGGCGCGCGCGCGCGCGCGCGCGGTC 120  
Db 4243 GCCGTACTCGGCCATTCTCCCGACCGGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 4302  
Qy 121 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACACCGCTGCAACCGCGCGCGCGCGCG 180  
Db 4303 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACACCGCTGCAACCGCGCGCGCGCG 4362  
Qy 181 CGGTTCGACTCGTGGCGCGGTCAACCTCGCAACACCGCTGCAACCGCGCGCGCGCGCG 240  
Db 4363 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACACCGCTGCAACCGCGCGCGCGCG 4422  
Qy 241 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACACCGCTGCAACCGCGCGCGCGCGCG 300  
Db 4423 CACGAGGGGTACCTCGCACCGCGCGGTGAGTCCGCAACACCGCTGCAACCGCGCGCG 4482  
Qy 301 CGGTTCGACTCGTGGCGCGGTCAACCTCGCAACACCGCTGCAACCGCGCGCGCGCGCG 360  
Db 4363 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACACCGCTGCAACCGCGCGCGCGCGCG 4422  
Qy 4423 CACGAGGGGTACCTCGCACCGCGCGGTGAGTCCGCAACACCGCTGCAACCGCGCGCG 4482

Qy 301 CGAGTCTCGCGCGGTGCGCGCTTCCCGCGGTTCAGGAGGGCGGCGTGGGACACGCTG 360  
Db 4483 CGGCGCTTGGCGGAATCGCCCTCGACCGGTGCGGGAGCGCGGGGTCTTCCGACACGCTC 4542  
Qy 361 CTGGCGCTCGCGCACACCGGACCGAACCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417  
Db 4543 CTGGCGCTCGCGCACACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4602  
Qy 418 GCGCGCGCGCGCGAGCGAGCGAAGTATGACGCGCATCTGACATCTCGGTCTGTCGCA 477  
Db 4603 GACCTTGTGCGGAGCGCGAGCGGTGATCGACGACCTGGAGCGCGGAGCGCTGATCGG 4662  
Qy 478 CGAGCCCTCGGGC 490  
Db 4663 ATGGCTCTCGGCC 4675

## RESULT 3

US-09-988-384B-34  
; Sequence 34, Application US/0988384B  
; Publication No. US20030073824A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.536US1  
; CURRENT APPLICATION NUMBER: US/09/988.384B  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: PCT/US99/14398  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 34  
; LENGTH: 4689  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-988-384B-34

Query Match 31.7%; Score 158.6; DB 10; Length 4689;  
Best Local Similarity 59.0%; Pred. No. 1.1e-27;  
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

Qy 1 GCGGTCCCGACGCGCGAGCAACCGCATCTCTGTAAGCTGGTCCGCGGCGCACTTCG 60  
Db 4183 GCGCTCCCGAGCGCGGCGCGCGGTCTCTCACTCCCTCGTCCGTACCCACGCGCG 4242  
Qy 61 ACGGTGCTCGGCCACAGCGGCGCGAGGCGATCGGCGCGCGCGCGCGCGCGCGGTC 120  
Db 4243 GCCGTACTCGGCCATTCTCCCGACCGGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 4302  
Qy 121 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACACCGCTGCAACCGCGCGCGCGCG 180  
Db 4303 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACACCGCTGCAACCGCGCGCGCGCG 4362  
Qy 181 CGGTTCGCGCGGAGCGTGTTCGATACCCACCGCGAGGCGGTGGTGGGTACTCTG 240  
Db 4363 AGGTTCGCGCGGAGCGTGTTCGATACCCACCGCGAGGCGGTGGTGGGTACTCTG 4422  
Qy 241 CGGTTCGAACTCTCTGCGGGAGGCGGACGAGGCTTGAACGGCGGGGGAAGACGACTCCGG 300  
Db 4423 CACGAGGGGTACCTTCGCAACCGCGCGGTGAGTCCGCAACACCGCTGCAACCGCGCG 4482  
Qy 301 CGAGTCTCGCGCGGTGCGCGCTTCCCGCGGTTCAGGAGGGCGGCGGTGGGACACGCTG 360  
Db 4483 CGGCGCTTGGCGGAATCGCCCTCGACCGGTGCGGGAGCGCGGGGTCTTCCGACACGCTC 4542  
Qy 361 CTGGCGCTCGCGCACACCGGACCGAACCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417  
Db 4543 CTGGCGCTCGCGCACACCGGACCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4602

418 GCGCGGCGCGGACGACGACGACTGACGCACTCCGCTCTCGTGCAA 477  
 QY  
 4603 GACCCCTGTGTGCGAGCGCGGGCTGATCGAGCACTCGGAGGCCCTGATCCGG 4662  
 Db  
 478 CGAGCCCTCGGGC 490  
 QY  
 4663 ATGGCTCTCGGCC 4675  
 Db

## RESULT 4

```

US-09-836-821-34
; Sequence 34, Application US/09835821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE INVENTION: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-34

```

Query Match 31.7%; Score 158.6; DB 10; Length 4689;  
Best Local Similarity 59.0%; Pred. No. 1.1e-27;  
Matches 291; Conservative 0; Mismatches 199; Indels 3;

Qy	1	GGCGTCCCGACGCCGCGAGCAGAACCGCATCTCTGCTGAAGCTGTGTCGCGCGCACGCTCG	60
Db	4183	GGCTCTCCCGAGCCCGAGCGCGGCGTCTCTCACTCGTCTCGATACCCACGCGGGCG	4242
Qy	61	ACGGTGTCTCGGCGCACACGGCGCGCGAAGGCATCGGCCCGCGCCAGGCGGTTCACAGGAGTC	120
Db	4243	GCCGTACTTGGGCCCATTCCTCCCCCGACCGGGTGGCCCCCGGCGGTGCTTACCGAGCTC	4302
Qy	121	GGCTTCGACTCGCTGGCGCGGCTCAACCTTCGCAAACAGCCTGCACGCGGCGCACCGGGCTG	180
Db	4303	GGCTTCGACTCGCTGAACGGCCGCTGACGTCGCAACAGCTCTCCACGGTGGTGGGCAAC	4362
Qy	181	CGGCTGCCCGCAGACGCTGATCTTCGACTTACCCACCCCGGAGGCGCTGGTCTGGCTACTG	240
Db	4363	AGGCTCCCGGCACCAACGGCTCTTCGACCAACGACGCGCGGCACTCGCGCGGCACTC	4422
Qy	241	CGGCTCGAATCTCTGCGGGAGCCGACGACGCGCTTCGACGGCGGGGGAAGACGACCTCGG	300
Db	4423	CACGAGGCGTACCTCGCACCGCGCGAGCGGCGCCCGACGGAATGGGAGGGCGGGTCCG	4482
Qy	301	CGAGTCTCTGCGGGCGCTCGGTTCTGCGCGGTTCAAGGAGGCGGCGCTGTGACACGCTG	360
Db	4483	CGGGCCCTTGGCCGAATCTGCCCTCTGACCGGCTTGGGGACGCGGGGGTCTCTGACACCGTC	4542
Qy	361	CTCGGCTCTCGCGCACACCGCGCACCGAAACGGGCGACGGACG---CCGAGACCCACCGAAGCG	417
Db	4543	CTGCGCTTCAACGGCATCGAGCCCGAGCCGAGTTCGGCGGTTTCGACGGCGGCGCCGCC	4602
Qy	418	GCCCGGCGCGCAGCAGCAGAACTGATCGACGCACTGGACATCTCGGCTCTCGTGCAA	477
Db	4603	GACCTCTGGTGGGAGCCGGAGCGTTCGATCGACGACCTGGACGCGCGAGGCCCTGATCCG	4662
Qy	478	CGAGCCCTCGGGC	490
Db	4663	ATGGCTCTCGGCC	4675

## RESULT 6

US-09-861-289-5

```

RESULT 5
US-10-271-889-34
; Sequence 34, Application US/10271889
; Publication No. US20030194784A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin
; FILE REFERENCE: 600.582U1
; CURRENT APPLICATION NUMBER: US/10/271.889
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-10-271-889-34

```

Query Match	31.7%;	Score 158.6;	DB 15;	Length 4689;
Best Local Similarity	59.0%;	Pred. No. 1.1e-27;		
Matches 291; Conservative	0;	Mismatches 199;	Indels	3;

[illegible]

## RESULT 6

US-09-861-289-5

```
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION: D.H.
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match      31.7%; Score 158.6; DB 9; Length 36778;
Best Local Similarity 59.0%; Pred. No. 6.8e-28;
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGGTCTCCCGACGCGGAGCAGAACCGCATCTCTGCTGAAGCTGTGTCGCGGCGCACGCTTCG 60
Db 31173 GCGGTCTCCCGAGCCCGAGCGCGCGCGCTCTCTACCTCTGTCCTGATCCACCGCGCGG 31232

QY 61 ACGGTGCTCGGCGCACAGCGCGCGGAGGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGG 120
Db 31233 GCGGTACTCGGCGCATCTCTCCCGGACCGGCTGCGCGCGCGCGCGCGCGCGCGCGCG 31292

QY 121 GGTTCGACTCGCTGCGCGCGGTCAACCTCCGCAACAGCCTGACGCGCGCGCGCGCGCGG 180
Db 31293 GGTTCGACTCGCTGCGCGCGGTGCGAGCTCGCGCAACAGCTCTCCACGGTGGTGGCAAC 31352

QY 181 CGGTGCTCGCGGAGCGGTGATCTTCGACTACCCACCGCGGAGGCGTGTGCGGTACTCG 240
Db 31353 AGGTCTCCCGCGCACCGAGCTTTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 31412

QY 241 CGGTGCAACTCTGCGGAGCGCGAGCGGCTGGAACGGCGGCGGAGGAGCGGCGGCGG 300
Db 31413 CACGAGCGGTACTCTGCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 31472

QY 301 CGAGTCTCTCGCGCGCGGTGCGCGGTTCGCGCGCGGTTCGAGGAGCGGCGCGGTGCTGG 360
Db 31473 CGGCGCTTGGCGGAACTGCGCGGTTCGCGCGGTGCGGAGCGGCGGCGGTCTCGACCG 31532

QY 361 CTGCGCTCTCGCGGAGCGCGGACCGGAAACCGGCGGCGGCGGCGGCGGCGGCGGCGG 417
Db 31533 CTGCGCTCTACCGGATCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 31592

QY 418 GCCCGCGCGCGCGAGCGAGCACTGATCGAGCGACTGAGCATCTCGGCTCTCGTGCAA 477
Db 31593 GACCTGTGCGGAGCGGAGCGGTGATCGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 31652

QY 478 CGAGCCCTCGGCG 490
Db 31653 ATGGCTCTCGGCC 31665

RESULT 7
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
```

```
; Sequence 5, Application US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match      31.7%; Score 158.6; DB 9; Length 36778;
Best Local Similarity 59.0%; Pred. No. 6.8e-28;
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGGTCTCCCGACGCGGAGCAGAACCGCATCTCTGCTGAAGCTGTGTCGCGGCGCACGCTTCG 60
Db 31173 GCGGTCTCCCGAGCCCGAGCGCGCGCGCTCTCTACCTCTGTCCTGATCCACCGCGCGG 31232

QY 61 ACGGTGCTCGGCGCACAGCGCGCGGAGGATCGGCGCGCGCGCGCGCGCGCGCGCGCGG 120
Db 31233 GCGGTACTCGGCGCATCTCTCCCGGACCGGCTGCGCGCGCGCGCGCGCGCGCGCGG 31292

QY 121 GGTTCGACTCGCTGCGCGCGGTCAACCTCCGCAACAGCCTGACGCGCGCGCGCGCGG 180
Db 31293 GGTTCGACTCGCTGCGCGCGGTGCGAGCTCGCGCAACAGCTCTCCACGGTGGTGGCAAC 31352

QY 181 CGGTGCTCGCGGAGCGGTGATCTTCGACTACCCACCGCGGAGGCGTGTGCGGTACTCG 240
Db 31353 AGGTCTCCCGCGCACCGAGCTTTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGG 31412

QY 241 CGGTGCAACTCTGCGGAGCGCGAGCGGCTGGAACGGCGGCGGAGGAGCGGCGGCGG 300
Db 31413 CACGAGCGGTACTCTGCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 31472

QY 301 CGAGTCTCTCGCGCGCGGTGCGCGGTTCGCGCGCGGTTCGAGGAGCGGCGCGGTGCTGG 360
Db 31473 CGGCGCTTGGCGGAACTGCGCGGTTCGCGCGGTGCGGAGCGGCGGCGGTCTCGACCG 31532

QY 361 CTGCGCTCTCGCGGAGCGCGGACCGGAAACCGGCGGCGGCGGCGGCGGCGGCGGCGG 417
Db 31533 CTGCGCTCTACCGGATCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 31592

QY 418 GCCCGCGCGCGCGAGCGAGCACTGATCGAGCGACTGAGCATCTCGGCTCTCGTGCAA 477
Db 31593 GACCTGTGCGGAGCGGAGCGGTGATCGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 31652

QY 478 CGAGCCCTCGGCG 490
Db 31653 ATGGCTCTCGGCC 31665

RESULT 8
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
```

```
; ORGANISM: Streptomyces venezuelae
US-09-836-821-5

Query Match      31.7%; Score 158.6; DB 10; Length 36778;
Best Local Similarity 59.0%; Pred. No. 6.8e-28;
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGTCCCGGAGCGGAGCAGACCGCATCTGCTGAAGCTGTCGCGGCGCACGCTTCG 60
Db 31173 GCGTCCCGGAGCGGAGCGGCGGCGGCGGCTCTCACCCTCGTCCGTACCCACGCGCG 31232

QY 61 ACGGTGCTCGGCGCACAGCGGCGCGGAGGCGATCGCGCGCGCGGCGGCTCCAGAGGTC 120
Db 31233 GCGTACTCGGCGCATCTCTCCCGGACCGGCTGCGCGCGGCGGCTTCCACCGAGCTC 31292

QY 121 GCGTTCGACTCGTGGCGCGGTCGAACCTCGCGCAACAGCGCTGCACCGGCGGCTG 180
Db 31293 GCGTTCGACTCGTGGCGCGGTCGAGCTTCGCAACAGCTCTCCACGGTGGTCGGCAAC 31352

QY 181 GCGTTCGACTCGTGGCGCGGTCGAACCTCGCGCAACAGCGCTGCACCGGCGGCTG 240
Db 31353 AGGCTCCCGCGCACCGGTCGAGCTTCGCAACAGCTCTCCACGGTGGTCGGCAAC 31412

QY 241 GCGTTCGACTCGTGGCGGAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 31413 CACGAGGCGGTACTCTCGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 31472

QY 301 CGAGTCTCTCGGCGCGTCCGCTTCCCGCGGTTCAAGAGGCGGCGGCGGCGGCGGCGG 360
Db 31473 CGGCGCTTGGCGGAACCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 31532

QY 361 CTCGCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417
Db 31533 CTGCGCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 31592

QY 418 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGA 477
Db 31593 GACCTCGTGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 31652

QY 478 CGAGCCCTCGGCGG 490
Db 31653 ATGGCTCTCGGCG 31665

RESULT 10
US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536U51
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
; LENGTH: 37948
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

Query Match      31.7%; Score 158.6; DB 10; Length 37948;
Best Local Similarity 59.0%; Pred. No. 6.8e-28;
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGGTCCCGGAGCGGCGGAGCAGAACCGCATCTGCTGAAGCTGTCGCGGCGCACGCTTCG 60
Db 32343 GCGTCCCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 32402
```





QY 121 GCTTCGACTCGCTGGCCGCGTCAACCTCCGAAACAGCCTGACGCGCCACCGGGCTG 180  
DB 29435 GCTTCGACTCGCTGACGCGCGTGCAGCTCCGCAACAGCTCTCCACGGTGTGCGCAAC 29494  
QY 181 CGGCTGCGCGGACGCTGATCTTGAATACCCACCGGAGGCGCTGGTGGCTACCTG 240  
DB 29495 AGGCTCCCGCGCACACGCTCTTGAACACCGCGCCGCGCACTCGCCGCGCACTC 29554  
QY 241 CGGCTGCAACTCTCGGAGGCGCGACGACGCGCTGGAAGCGGCGGGAAGACGACTCCG 300  
DB 29555 CACGAGGCGTACTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29614  
QY 301 CGAGTCTCTCGGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 360  
DB 29615 CGGCGCTTGGCGCGACTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 29674  
QY 361 CTCGGCTTGGCGCGACCGCGCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 417  
DB 29675 CTGCGCTTACCGCGATCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29734  
QY 418 GCCCG 477  
DB 29735 GACCTTGTGCGGAGCGCGAGCGCGCTGATCGACGCTGGAACCTGGAACCGCG 29794  
QY 478 CGAGCCTCTCGGC 490  
DB 29795 ATGCTCTCGGC 29807

## RESULT 13

US-10-160-539-19  
; Sequence 19, Application US/10160539  
; Publication No. US20030162262A1  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120  
; CURRENT APPLICATION NUMBER: US/10/160,539  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: US/09/657,440  
; PRIOR FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: 09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-10-160-539-19

Query Match 31.7%; Score 158.6; DB 15; Length 38506;  
Best Local Similarity 59.0%; Pred. No. 6.7e-28;  
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;  
QY 1 GCGGTCCCGACCGCGGAGCAGAACCGCATCTCTGTAAGCTGTCCGCGCGCCACGCTTCG 60  
DB 29315 GCGTCCCGGAGCG 29374  
QY 61 AGGTGCTCGGCGCAGCGCGCGCGAGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
DB 29375 GCGTACTCGGCGCATCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29434  
QY 121 GCTTCGACTCGCTGCGCGCGGTCAACTCCGCAACAGCTGCAACGCGCGCGCGCGCG 180  
DB 29435 GCTTCGACTCGCTGACGCGCGGTGAGCTCGCGCAACAGCTCTCCACGCTGTGCGCAAC 29494

QY 181 CGGCTGCGCGGAGCCTGATCTTGAATACCCACCGCGAGCGCTGGTGGCTACCTG 240  
DB 29495 AGGCTCCCGCGCACACGCTCTTGAACACCGCGCCGCGCACTCGCCGCGCACTC 29554  
QY 241 CGGCTGCAACTCTCGGAGGCGCGACGACGCGCTGGAAGCGGCGGGAAGACGACTCCG 300  
DB 29555 CACGAGGCGTACTCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29614  
QY 301 CGAGTCTCTCGGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 360  
DB 29615 CGGCGCTTGGCGCGACTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 29674  
QY 361 CTCGGCTTGGCGCGACCGCGCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 417  
DB 29675 CTGCGCTTACCGCGATCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29734  
QY 418 GCCCG 477  
DB 29735 GACCTTGTGCGGAGCGCGAGCGCGCTGATCGACGCTGGAACCTGGAACCGCG 29794  
QY 478 CGAGCCTCTCGGC 490  
DB 29795 ATGCTCTCGGC 29807

## RESULT 14

US-10-205-032-17  
; Sequence 17, Application US/10205032  
; Publication No. US20030113874A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Yang, Xianshu  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN  
; FILE REFERENCE: 3016-2US  
; CURRENT APPLICATION NUMBER: US/10/205,032  
; CURRENT FILING DATE: 2002-07-26  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 17  
; LENGTH: 4725  
; TYPE: DNA  
; ORGANISM: micromonospora carbonacea subspecies aurantiaca  
US-10-205-032-17

Query Match 31.1%; Score 155.8; DB 15; Length 4725;  
Best Local Similarity 62.2%; Pred. No. 4.7e-27;  
Matches 263; Conservative 0; Mismatches 157; Indels 3; Gaps 1;  
QY 12 CGCGGAGCAGAACCGCATCTCTGTAAGCTGTGTCGCGCGCACGCTTCGACGCTGCTCGG 71  
DB 4254 CGACGAAACCGCGCGCGCGCTGCTCGACCTGTATACCGCGGAGTCCCGCGCTCTCGG 4313  
QY 72 CCACAGCGCGCGCGAAGGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 131  
DB 4314 CCACCGCGCGCGCGAGCAGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 4373  
QY 132 GTGCGCGCGCGTCAACCTCCGCAACAGCTGCAACGCGCGCGCGCGCGCGCGCGCGCG 191  
DB 4374 GTGACCGCGCGTCAACCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4433  
QY 192 GACGCTGATCTTCGACTACCCACCGCGAGCGCTGTGTCGCTACCTACCTGCGGTGCACT 251  
DB 4434 CACCTCTGCTTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4493  
QY 252 CTTGCGGAGCGCGCAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 311  
DB 4494 CGGTCCCGCT 4550  
QY 312 GCGCGCGCGTTCGCGCGGTTCAGGAGCGCGCGCGGTGCTGGAACGCTGCTGCGCTGCG 371  
DB 4551 GCACCTGCGCTGCGCGCGGTTCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 4610



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:18:23 ; Search time 1960 Seconds  
(without alignments)  
7633.139 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543  
Perfect score: 501  
Sequence: 1 gcgggtcccgacgcgagca.....ccctcgggcagagagctga 501

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	75	15.0	925	29	CNS0091P
C 2	74	14.8	924	13	AX442207
C 3	72.8	14.5	924	13	AX442207
C 4	71.4	14.3	982	13	AX415111

5	71.2	14.2	925	29	CNS0091P
c	71	14.2	935	29	CNS006XK
c	70.8	14.1	1104	13	BQ722521
8	70	14.0	1232	29	AG072425
c	68.8	13.7	803	29	AG042898
10	68.8	13.7	932	29	AG042898
c	68.6	13.7	932	29	CNS0072Q
12	67.6	13.5	935	29	CNS0072Q
13	66.8	13.3	622	14	CNS006XK
14	66.4	13.3	1516	12	CB654160
15	66.2	13.2	562	13	BQ606671
16	66	13.2	621	14	CF326124
17	66	13.2	633	14	CF326834
18	65.8	13.1	982	13	EX415111
19	65.8	13.1	1134	12	EM915656
c	63.8	12.7	965	28	AQ893417
21	63.2	12.6	438	13	BQ805019
22	63	12.6	710	9	AU163439
c	63	12.6	1144	13	EX415926
24	62.6	12.5	638	28	BZ895164
c	62.2	12.4	1094	29	AG036373
26	61.8	12.3	449	10	BF728957
27	61.8	12.3	790	14	CB671171
28	61.8	12.3	829	14	CB642787
29	61.6	12.3	997	29	AG057392
c	61.4	12.3	805	29	CG287915
31	61.4	12.3	840	29	CG327544
32	61.2	12.2	615	14	CB627718
33	61.2	12.2	688	29	CG263720
34	61.2	12.2	771	14	CB670011
35	61.2	12.2	1101	29	CNS0175X
36	61	12.2	653	13	BU673539
37	60.8	12.1	718	14	CD868844
c	60.8	12.1	832	29	CC719244
c	60.8	12.1	841	29	CG291746
40	60.8	12.1	918	29	CC719255
41	60.8	12.1	959	29	CG341580
42	60.8	12.1	1112	14	CA477644
c	60.8	12.1	1122	12	BM914385
44	60.6	12.1	738	14	CF868028
45	60.6	12.1	810	14	CB898085

#### ALIGNMENTS

RESULT 1  
CNS0091P/c  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION  
AL053013  
VERSION  
AL053013.1  
KEYWORDS  
GSS.  
SOURCE  
Drosophila melanogaster (fruit fly)  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 925)  
Genoscope.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
JOURNAL  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosagawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of

CNS0091P 925 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL053013 GI:4934461

GSS.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 925)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

JOURNAL

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Oosagawa and

Aaron Mammosier in Pieter de Jong's laboratory in the Department of









**AUTHORS**  
**TITLE** NTH-MGC <http://mgc.nci.nih.gov/>.  
**JOURNAL** National Institutes of Health, Mammalian Gene Collection (MGC)  
**COMMENT** Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-x@mail.nih.gov](mailto:cgapbs-x@mail.nih.gov)  
 Tissue Procurement: Dr. James R. Lupski  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LMAM13583 row: 9 column: 03  
 High quality sequence stop: 297.

**FEATURES**  
 source  
 1. .1104  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6188186"  
 /sex="male"  
 /tissue\_type="sympathetic trunk"  
 /dev\_stage="adult, 16 yr"  
 /lab\_host="DH10B"  
 /clone\_lib="Lupski sympathetic trunk"  
 /note="vector: pCMV-SPORT6 (Life Technologies); Site\_1: NotI; Site\_2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TCGACCCACCGCTCCG-3' and 5'-GACTAGTTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."

## ORIGIN

Query Match 14.1%; Score 70.8; DB 13; Length 1104;  
 Best Local Similarity 47.0%; Pred. No. 3.2;  
 Matches 215; Conservative 0; Mismatches 241; Indels 1; Gaps 1;

Qy 39 GCTGGTCCGGGCCACGCTTCGACGGTGTCTGGCCACAGCGCGCCGCGAAGGATCGGCC 98  
 |||  
 Db 1069 GCGGACGCGCCCGNCCCGGNGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1010

Qy 99 GCGCCAGCGGTTTCAAGAGTTCGGCTTCGACTCGCTGGCGCGCGGTCAACTTCGCGAAG 158  
 |||  
 Db 1009 GCCCGCGGGCTCG 950

Qy 159 CTTGACGCGCGCCACCGGGCTCGCGCTGCGCGCGACGCTGATCTTCGACTACCCACCCC 218  
 |||  
 Db 949 GGCNCGNCCGCGCCCG 890

Qy 219 GGAGGCGCTGTCTGGCTTACTTCGCGCTGCAACTCTCTGGGAGGCGGACGACGCGCTTGA 278  
 |||  
 Db 889 CGCGGGCG 830

Qy 279 CGGCGCGGGAAGAAGACTTCGCGGAGTCTCTCGCGCGCGGTGCGCTTCGCGCGGTTCAAG 338  
 |||  
 Db 829 CGCCG 770

Qy 339 GGGGGGCTGTGACACGCTCTCGGCTCGCGGACGACCGGACCGGACCGGACCGGACCGG 398  
 |||  
 Db 769 CCCCGCGGGCG 710

Qy 399 CGCGGAGACACCGAAGCG 458  
 |||  
 Db 709 CCCNCCCG 495

Qy 459 CATCTTCGGTTCGTGCAACGAGCCCTCGGGCAGACG 495  
 |||  
 Db 649 CGGC-CGCGGCGGCG 614

## RESULT 8

AG072425  
 LOCUS Pan troglodytes DNA, clone: PTB-063017.F, genomic survey sequence.  
 AG072425  
 DEFINITION AG072425  
 ACCESSION AG072425  
 VERSION AG072425.1 GI:16624227  
 KEYWORDS GSS.  
 SOURCE Pan troglodytes (chimpanzee)  
 ORGANISM Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 1  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
 BAC end sequences of Library PTB  
 Unpublished  
 2 (bases 1 to 1232)  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
 Direct Submission  
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: [chimpanzee@gsc.riken.go.jp](mailto:chimpanzee@gsc.riken.go.jp), URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.  
 PRIMERS  
 Sequencing: -21M13  
 LIBRARY  
 Vector : pXS145  
 R.Site 1 : SacI  
 R.Site 2 : SacI.  
 Location/Qualifiers  
 1. .1232  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="PTB-063017.F"  
 /sex="male"  
 /cell\_type="lymphoblast"  
 /clone\_lib="PTB Chimpanzee Male BAC Library"

**FEATURES**  
 source  
 1. .1232  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="PTB-063017.F"  
 /sex="male"  
 /cell\_type="lymphoblast"  
 /clone\_lib="PTB Chimpanzee Male BAC Library"

**ORIGIN**  
 Query Match 14.0%; Score 70; DB 29; Length 1232;  
 Best Local Similarity 47.1%; Pred. No. 4.1;  
 Matches 210; Conservative 0; Mismatches 235; Indels 1; Gaps 1;

Qy 42 GGTCCGCGGCACGCTTCGACGGTGTCTCGGCACAGCGGGCGCCGAGGATCGGCCCGCG 101  
 |||  
 Db 535 GGGCG 594

Qy 102 CCAGGCGTTCAGAGGTTCGCTTCGACTCGCTGGCGCGGTCAACTTCGCGCAACAGCCT 161  
 |||  
 Db 595 CGGCGGCG 654

Qy 162 GCACGCGCGCCACCGGGCTGCGGCTGCGCCGCGAGCGTCTTCGACTACCCACCCCGGA 221  
 |||  
 Db 655 GCCCG 714

Qy 222 GGCCTGCTGCTGCTACTTCGCGCTGCGAACTCTTCGCGGAGCGCCGACGCGGCTTGAGCG 281  
 |||  
 Db 715 CGCGNGCG 774

Qy 282 GCGGGAAGACGACCTTCGCGGAGTCTTCGCGCGCGGTGCGCTTCGCGGTTCAAGGAGCG 341  
 |||  
 Db 775 GCG 834

Qy 342 GGGCGTGTGGACACGCTGTCTCGGCTTCGCGGACACCGGACACCGGACCGGACGCGCG 401  
 |||  
 Db 835 GCGCGGCG 893







ORIGIN

Query Match 13.3%; Score 66.4; DB 12; Length 1516;  
Best Local Similarity 48.0%; Pred. No. 13;  
Matches 182; Conservative 0; Mismatches 194; Indels 3; Gaps 1;  
QY 64 GTGCTGGCCACAGCGGCGCCGAAGGCATCGGCCCGCCGACGCGCTTCCAGAGGTTCGCG 123  
Db 357 GCGCGCGGGCGCGCGCCCGCCGCGGNCNCCGCGGCGCGGCGCGCGCGCGCGCGCGCG 416  
QY 124 TTCTGACTCTGCTGCGCGCTCAACTCTCGCAACAGAGCTGCAAGCGGCGCAACCGGGTTCGCG 183  
Db 417 NCGGCG 476  
QY 184 CTGCG 243  
Db 477 CCG 536  
QY 244 GTGCAACTCTGCGGAGCGCGAGCGCGCTGGAAGCGGCGGGAAGACGACCTTCGCGCGA 303  
Db 537 GCG 596  
QY 304 GTCTCTGCG 363  
Db 597 GCG 656  
QY 364 GCGCTCG 420  
Db 657 GCG 716  
QY 421 CCG 439  
Db 717 GCG 735

RESULT 15

BQ606671 562 bp mRNA linear EST 25-JUN-2002  
LOCUS BRY 2539 wheat EST endosperm library Triticum aestivum cDNA 5',  
DEFINITION mRNA sequence.

ACCESSION BQ606671  
VERSION BQ606671.1 GI:21555998

KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum

REFERENCE 1. (bases 1 to 562)  
AUTHORS Clarke, B., Lambrecht, M. and Rhee, S.Y.  
TITLE Arabidopsis genomic information for interpreting wheat EST sequences

JOURNAL Arabidopsis  
MEDLINE Arabidopsis  
COMMENT Arabidopsis genomic information for interpreting wheat EST sequences

CONTACT: Lambrecht M  
The Arabidopsis Information Resource  
Carnegie Institution of Washington, Dept. of Plant Biology  
260 Panama Street, Stanford, CA 94305, USA  
Tel: 1 650 325 1521 x 251  
Fax: 1 650 325 3748  
Email: rhee@acoma.stanford.edu.

Location/Qualifiers

1. .562  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Wyuana"  
/db\_xref="taxon:4565"  
/tissue\_type="endosperm"  
/dev\_stage="developing endosperm tissue 8, 10 and 12 DPA  
(days post anthesis)"  
/clone\_lib="wheat EST endosperm library"

FEATURES

source

1. .562

/organism="Triticum aestivum"

/mol\_type="mRNA"

/cultivar="Wyuana"

/db\_xref="taxon:4565"

/tissue\_type="endosperm"

/dev\_stage="developing endosperm tissue 8, 10 and 12 DPA  
(days post anthesis)"

/clone\_lib="wheat EST endosperm library"

ORIGIN

Query Match 13.2%; Score 66.2; DB 13; Length 562;  
Best Local Similarity 50.8%; Pred. No. 12;  
Matches 158; Conservative 0; Mismatches 153; Indels 0; Gaps 0;  
QY 127 GACTCTGGCGCGCGGTCAACCTCCGCAACAGCCTGCAAGCGGCGCAACCGGGGTGCGGCTG 186  
Db 132 GCGCGGTGGCGCGGTCAAGAGCGGCGAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 191  
QY 187 CCGGGAAGCGCTGATCTTCTGACTACCCACCGCGGCGCGCTGCTGCTGCTGCTGCTGCTG 246  
Db 192 TTCTGAGAGCGTGGCGCTCATCTCGGCTCCACGCGGATCGTCTGCGACCTCTCTCTCTCGAC 251  
QY 247 GAACTCTCTGCGGAGGCGCGAGCGAGCGGCTTGAAGCGGCGGGAAGACGACCTTCGCGCGAGTC 306  
Db 252 ATCTCTCCCGGGAAGACACCGCGGCGGCGCTTGAAGGTGTACCGCGTCTCCCGCGCG 311  
QY 307 CTGCGCGCGCGTGGCGCTTCTGCGCGCGGTTCAGAGAGCGGCGGTGCTGAGACAGCTGCTGCGC 366  
Db 312 GCGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 371  
QY 367 CTGCG 426  
Db 372 CTGCG 431  
QY 427 GCGGAGCGCGC 437  
Db 432 TTCTAGCGCGC 442

Search completed: June 17, 2004, 03:18:54  
Job time : 1964 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 16, 2004, 15:37:20 ; Search time 2102.33 Seconds  
(without alignments)  
7116.356 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325  
Perfect score: 501  
Sequence: 1 atgttctacagtcgggcac.....ccgcgctcagggctcgtgc 501

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 15

Total number of hits satisfying chosen parameters: 13643

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pin:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rtd:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gsa1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
1	21	4.2	303 10	BF896427
2	21	4.2	316 10	BF896426
3	20	4.0	546 28	BH516834
4	20	4.0	709 28	BH585470

115	9	AA065382	115	9	AA065382
249	9	AA492813	249	9	AA492813
324	9	AA881756	324	9	AA881756
378	10	BE418061	378	10	BE418061
378	10	BE418061	378	10	BE418061
420	12	BG837626	420	12	BG837626
431	10	BF411078	431	10	BF411078
438	14	CB288581	438	14	CB288581
445	9	A1746181	445	9	A1746181
446	9	A1833600	446	9	A1833600
449	14	CA402106	449	14	CA402106
469	12	BG267470	469	12	BG267470
474	14	CF608321	474	14	CF608321
476	14	CF608320	476	14	CF608320
485	12	BG265809	485	12	BG265809
507	14	CB863203	507	14	CB863203
522	9	A1737095	522	9	A1737095
528	9	A1737056	528	9	A1737056
530	14	CA401348	530	14	CA401348
538	14	CF486345	538	14	CF486345
541	9	A1677352	541	9	A1677352
543	14	CA404459	543	14	CA404459
544	14	CF488024	544	14	CF488024
547	14	CB288531	547	14	CB288531
562	14	CB288483	562	14	CB288483
567	14	CA398780	567	14	CA398780
571	9	A1714654	571	9	A1714654
578	28	CC026191	578	28	CC026191
581	14	CF489087	581	14	CF489087
583	14	CB288474	583	14	CB288474
584	12	BG266929	584	12	BG266929
586	9	A1711641	586	9	A1711641
590	9	A1657469	590	9	A1657469
601	14	CD444963	601	14	CD444963
610	9	A1759026	610	9	A1759026
616	14	CB919676	616	14	CB919676
617	9	A1745808	617	9	A1745808
617	9	AI833967	617	9	AI833967
621	12	BM356065	621	12	BM356065
625	14	CD428301	625	14	CD428301
631	14	CD442989	631	14	CD442989
636	14	CA400622	636	14	CA400622

ALIGNMENTS

RESULT 1

BF896427

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BF896427 303 bp mRNA linear EST 18-JAN-2001  
RCL-WT0147-241100-011-f10 MT0147 Homo sapiens cDNA, mRNA sequence.

BF896427 GI:12287886

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 303)

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,

O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

10737800

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Frudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Thu Jun 17 09:02:23 2004

Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1&t2=RC1-MT0147-241100-011-f09&t3=2000-11-24&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 24  
High quality sequence stop: 303.  
Location/Qualifiers  
1. .303  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="MT0147"  
/note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES  
source  
Query Match 4.2%; Score 21; DB 10; Length 303;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN  
QY 62 GCCTCTGGTCGCTCGCTTCCT 82  
Db 134 GCCTCTGGTCGCTCGCTTCCT 154

RESULT 2  
BF896426 316 bp mRNA linear EST 18-JAN-2001  
LOCUS RCL-MT0147-241100-011-f09 MT0147 Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
ACCESSION BF896426  
VERSION BF896426.1 GI:12287885  
KEYWORDS  
SOURCE EST.  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 316)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
PUBMED 10737800  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1&t2=RC1-MT0147-241100-011-f09&t3=2000-11-24&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 14

High quality sequence stop: 316.  
Location/Qualifiers  
1. .316  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="MT0147"  
/note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES  
source  
Query Match 4.2%; Score 21; DB 10; Length 316;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN  
QY 62 GCCTCTGGTCGCTCGCTTCCT 82  
Db 147 GCCTCTGGTCGCTCGCTTCCT 167

RESULT 3  
BH516834/c 546 bp DNA linear GSS 13-DEC-2001  
LOCUS BOGG50TF BOGG Brassica oleracea genomic clone BOGG50, genomic survey sequence.  
DEFINITION  
ACCESSION BH516834  
VERSION BH516834.1 GI:17724924  
KEYWORDS  
SOURCE GSS.  
Brassica oleracea  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 546)  
Town, C.D., Van Aken, S., Utterback, T., Koc, H. and Fraser, C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOGG50TR  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: sheared ends.  
Location/Qualifiers  
1. .546  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO100DH3"  
/db\_xref="taxon:3712"  
/clone\_lib="BOGG50"  
/notes="Vector: pHSI; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHSI using BstXI linkers"

FEATURES  
source  
Query Match 4.0%; Score 20; DB 28; Length 546;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN  
QY 226 GACGACCTGATGCGGCTGAT 245  
Db 399 GACGACCTGATGCGGCTGAT 380

Thu Jun 17 09:02:23 2004

RESULT 4  
BH585470/c  
LOCUS  
DEFINITION  
BH585470 BOGE Brassica oleracea genomic clone BOGEK50, genomic survey sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Brassica oleracea  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE  
1 (bases 1 to 709)  
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M. Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other\_GSSs: BOGEK50TF  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers  
1..709  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOGEK50"  
/clone\_lib="BOGE"  
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

FEATURES  
source  
Query Match 4.0%; Score 20; DB 28; Length 709;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 226 GACGACGTGATCGGCTGAT 245  
Db 431 GACGACGTGATCGGCTGAT 412

RESULT 5  
AA065382  
LOCUS  
DEFINITION  
AA065382 115 bp mRNA linear EST 03-FEB-1997  
clone IMAGE:515552 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 115)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

FEATURES  
source  
Query Match 4.0%; Score 20; DB 28; Length 709;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 226 GACGACGTGATCGGCTGAT 245  
Db 431 GACGACGTGATCGGCTGAT 412

Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:309400  
Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 88.  
Location/Qualifiers  
1..115  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CD-1"  
/db\_xref="taxon:10090"  
/clone="IMAGE:515552"  
/sex="males"  
/tissue\_type="testis"  
/dev\_stage="10-12 week old"  
/lab\_host="SOLR (kanamycin resistant)"  
/clone\_lib="Stratagene mouse testis (#937308)"  
/note="Organ: testis; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGACG 3' -3' adaptor sequence: 5' CTCGAGTTT TTT TTT TTT TTT 3' "

ORIGIN  
Query Match 3.8%; Score 19; DB 9; Length 115;  
Best Local Similarity 100.0%; Pred. No. 8.6e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 267 CTGCGCGCGGCGCGGACC 285  
Db 21 CTGCGCGCGGCGGCGGACC 39

RESULT 6  
AA492813  
LOCUS  
DEFINITION  
AA492813 249 bp mRNA linear EST 25-JUN-1997  
clone IMAGE:918391 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 249)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:530607  
Seq primer: -28ml3 rev1 ET from Amersham.  
Location/Qualifiers  
1..249  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CD-1"  
/db\_xref="taxon:10090"

FEATURES  
source  
Query Match 3.8%; Score 19; DB 9; Length 115;  
Best Local Similarity 100.0%; Pred. No. 8.6e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 267 CTGCGCGCGGCGGACC 285  
Db 21 CTGCGCGCGGCGGCGGACC 39

RESULT 6  
AA492813  
LOCUS  
DEFINITION  
AA492813 249 bp mRNA linear EST 25-JUN-1997  
clone IMAGE:918391 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 249)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:530607  
Seq primer: -28ml3 rev1 ET from Amersham.  
Location/Qualifiers  
1..249  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CD-1"  
/db\_xref="taxon:10090"

Thu Jun 17 09:02:23 2004

adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

```

/clone="IMAGE:918391"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse testis (#937308)"
/notes="Organ: testis; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN
Query Match          3.8%; Score 19; DB 9; Length 249;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 CCTGGCGCGGTCCGACC 285
Db 144 CCTGGCGCGGTCCGACC 162

RESULT 7
AA881756          324 bp mRNA linear EST 26-MAR-1998
LOCUS             vx27b02.r1 Soares mammary_gland NbMWG Mus musculus cDNA clone
DEFINITION        IMAGE:1265643 5' similar to SW:FNPP_SCHPO Q00472
4-NITROPHENYLPHOSPHATASE 1, mRNA sequence.
ACCESSION         AA881756
VERSION           AA881756.1 GI:2991066
KEYWORDS          Mus musculus (house mouse)
SOURCE            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ORGANISM          1 (bases 1 to 324)
REFERENCE         Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE             The WashU-HMI Mouse EST Project
JOURNAL           Unpublished (1996)
COMMENT           Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:668195
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 308.
Location/Qualifiers
1..324
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1265643"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares mammary gland NbMWG"
/notes="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAGTGGCGCGCGCAATGTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI

us-10-042-665a-3_copy_1825_2325.ol.rst

adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."
```

Thu Jun 17 09:02:23 2004

```

MEDLINE
PUBMED
COMMENT
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Oligo-dt track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=No.
Location/Qualifiers
1. .431
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CN0-bmg-e-12-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CN0"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-CN0
library is a normalized library constructed from the
following tissues: brown adipose, penis, salivary glands,
bladder, fundus, cervix, seminal vesicles. For a detailed
description of the library from which this clone was
derived, please visit our web site at
raster.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldi, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_SEQ=None found"

ORIGIN
Query Match 3.8%; Score 19; DB 10; Length 431;
Best Local Similarity 100.0%; Pred. No. 9.5e+02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 360 GGGCGCGGTCTCCGCGCC 378
|||
Db 335 GGGCGCGGTCTCCGCGCC 353

RESULT 11
CB288561/C
LOCUS
DEFINITION
V-B-16F09 VAN-Baker-1 Vitis aestivalis cDNA clone V-B-16F09 5',
mRNA sequence.
ACCESSION
CB288581
VERSION
CB288581.1 GI:28602322
KEYWORDS
EST.
SOURCE
Vitis aestivalis
ORGANISM
Vitis aestivalis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; Vitaceae; Vitis.
1 (bases 1 to 438)
Hou, H.S., Phanikant, T.V., Kovacs, L. and Qiu, W.P.
Expressed sequence tags of young leaf tissues of a
disease-resistant Vitis aestivalis var. Norton
Unpublished (2003)
JOURNAL
Contact: Wenping Qiu
Department of Fruit Science
Southwest Missouri State University-Mountain Grove
9740 Red Spring Road, Mountain Grove, MO 65711-2999, USA
Tel: 417 926 4105
Fax: 417 926 6646
Email: weq070f@msu.edu
Insert Length: 438 Std Error: 0.00

mays cDNA clone Zm10_04b04, mRNA sequence.
BG837626
BG837626.1 GI:14203949
VERSION
KEYWORDS
EST.
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 420)
Harris, L.J., Balcerzak, M., Allard, S., Saparno, A., Couroux, P., De
Moors, A., Hattori, J.I., Ouellet, T., Robert, L.S., Singh, J.A.,
Spott, D. and Tinker, N.A.
Expressed Sequence Tags from Maize Silk Six Hours After Silk
Channel Inoculation with Fusarium graminearum
Unpublished (2001)
JOURNAL
COMMENT
Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harris.lj@em.agr.ca.
Location/Qualifiers
1. .420
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="CO388"
/db_xref="taxon:4577"
/clone="Zm10_04b04"
/tissue_type="Silk"
/dev_stage="4-5 days post-silk emergence"
/clone_lib="Zm10_AAFEC_EOCC_Fusarium_graminearum_corn_silk"
/note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;
Site 2: XhoI; Field-grown corn was silk channel-inoculated
in the morning (~10 am) with 1 ml of a macroconidial
suspension (500,000 spores/ml) of Fusarium graminearum and
silk channels were collected and immediately frozen in
liquid nitrogen 6 hours later. RNA was extracted from
silk tissue between 1 cm below and above the inoculation
point in the silk channel, RNA from five silk channels was
pooled."

ORIGIN
Query Match 3.8%; Score 19; DB 12; Length 420;
Best Local Similarity 100.0%; Pred. No. 9.5e+02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 125 TCTGGCGGTCTCCGCTGTT 143
|||
Db 204 TCTGGCGGTCTCCGCTGTT 186

RESULT 10
BF411078
LOCUS
DEFINITION
UI-R-CN0-bmg-e-12-0-UI.s1 UI-R-CN0 Rattus norvegicus cDNA clone
UI-R-CN0-bmg-e-12-0-UI 3', mRNA sequence.
ACCESSION
BF411078
VERSION
BF411078.1 GI:11399058
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 431)
Bonaldi, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

Thu Jun 17 09:02:23 2004

Plate: VAN-Baker-1-6 row: F column: 09

Seq primer: T3 PRIMER

High quality sequence stop: 438

POLYA=NO. Location/Qualifiers

FEATURES  
source

1. .438  
 /organism="Vitis aestivalis"  
 /mol\_type="mRNA"  
 /cultivar="Norton"  
 /db\_xref="taxon:3605"  
 /clone="V-B-16F09"  
 /tissue\_type="leaf"  
 /dev\_stage="Young leaf"  
 /lab\_host="XL10-Gold E. coli"  
 /clone\_lib="VAN-Baker-1"  
 /note="Vector: pBluescript II SK (+) Phagemid; Site 1: Xho I; Site 2: EcoRI; VAN-Baker-1 is a cDNA library of Norton grape young leaves (Vitis aestivalis var. Norton) grapevines were grown under normal greenhouse conditions. The cDNA synthesis and library construction was performed according to the instruction manual for pBluescript II XR cDNA library construction kit provided by Stratagene."

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 438;  
 Best Local Similarity 100.0%; Pred. NO. 9.5e+02; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0;

QY 221 CGCGGACGACGTGATGCG 239

Db 140 CGCGGACGACGTGATGCG 122

## RESULT 12

AI746181 445 bp mRNA linear EST 02-FEB-2000  
 LOCUS 605082C07.x1 605 - EndospERM cDNA library from Schmidt lab Zea mays  
 DEFINITION cDNA, mRNA sequence.

ACCESSION AI746181 GI:5124445

VERSION EST.

KEYWORDS Zea mays

SOURCE Zea mays

ORGANISM  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 445)

REFERENCE

AUTHORS Walbot, V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL University

COMMENT Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 605082 row: C column: 07.

## FEATURES

source

1. .445  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /cultivar="Ohio43"  
 /db\_xref="taxon:4577"  
 /tissue\_type="nucellar, embryo, and endospERM"  
 /dev\_stage="10-14 days post-pollination"  
 /lab\_host="DH5 (alpha)"  
 /clone\_lib="605 - EndospERM cDNA library from Schmidt lab"  
 /note="Organ: Kernel; Vector: pAD-GAL4-2; Site 1: EcoRI;  
 Site 2: XhoI; Kernel endospERM cDNA library from Schmidt lab"

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 445;

Best Local Similarity 100.0%; Pred. NO. 9.5e+02; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0;

QY 271 GCGGCGGTGCGGACCACT 289

Db 366 GCGGCGGTGCGGACCACT 384

## RESULT 13

AI833600 446 bp mRNA linear EST 02-FEB-2000  
 LOCUS 605092F11.x1 605 - EndospERM cDNA library from Schmidt lab Zea mays  
 DEFINITION cDNA, mRNA sequence.

ACCESSION AI833600

VERSION AI833600.1 GI:5455910

KEYWORDS EST.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 446)

REFERENCE

AUTHORS Walbot, V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL University

COMMENT Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 605092 row: F column: 11.

## FEATURES

source

1. .446  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /cultivar="Ohio43"  
 /db\_xref="taxon:4577"  
 /tissue\_type="nucellar, embryo, and endospERM"  
 /dev\_stage="10-14 days post-pollination"  
 /lab\_host="DH5 (alpha)"  
 /clone\_lib="605 - EndospERM cDNA library from Schmidt lab"  
 /note="Organ: Kernel; Vector: pAD-GAL4-2; Site 1: EcoRI;  
 Site 2: XhoI; Kernel endospERM cDNA library from Schmidt lab"

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 446;

Best Local Similarity 100.0%; Pred. NO. 9.5e+02; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0;

QY 271 GCGGCGGTGCGGACCACT 289

Db 374 GCGGCGGTGCGGACCACT 392

## RESULT 14

CA402106 449 bp mRNA linear EST 07-NOV-2002  
 LOCUS ELO1N0431E02.9 EndospERM\_4 Zea mays cDNA, mRNA sequence.

ACCESSION CA402106

VERSION CA402106.1 GI:24766961

KEYWORDS EST.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 449)



Thu Jun 17 09:02:23 2004

**AUTHORS** Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and Messing, J.  
**TITLE** Sequencing of the maize endosperm ESTs  
**JOURNAL** Unpublished (2002)  
**COMMENT** Contact: Lai, Jinsheng  
 Dr. Joachim Messing's lab  
 Waksman Institute, Rutgers University  
 190 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
 Tel: 732-445-3801  
 Fax: 732-445-5735  
 Email: jlai@waksman.rutgers.edu  
**Seq primer:** T7.

**FEATURES** source  
 1. .449  
 Location/Qualifiers  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /cultivar="W22"  
 /db\_xref="taxon:4577"  
 /tissue\_type="Endosperm of 7-23DAP"  
 /clone\_lib="Endosperm 4"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

**ORIGIN**  
 Query Match 3.8%; Score 19; DB 14; Length 449;  
 Best Local Similarity 100.0%; Pred. No. 9.5e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 271 GCCGGCGTGGCGACCACT 289  
 |||||  
**Db** 411 GCCGGCGTGGCGACCACT 429

**RESULT 15**  
**LOCUS** BG267470 469 bp mRNA linear EST 20-FEB-2001  
**DEFINITION** 100122C06.x1 1000 - Unigene 1 from Maize Genome Project Zea mays cDNA, mRNA sequence.

**ACCESSION** BG267470  
**VERSION** BG267470.1 GI:12971414  
**KEYWORDS** EST.  
**SOURCE** Zea mays  
**ORGANISM** Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

**REFERENCE** 1 (bases 1 to 469)  
**AUTHORS** Walbot, V.  
**TITLE** Maize ESTs from various cDNA libraries sequenced at Stanford University  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu

**FEATURES** source  
 1. .469  
 Location/Qualifiers  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /db\_xref="dbEST:605097G06.x1"  
 /db\_xref="taxon:4577"

/clone\_lib="1000 - Unigene 1 from Maize Genome Project"  
 /note="This library represents the unique ESTs found in the first round of EST sequencing at Stanford University for the maize genome project. Sequences are present from libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683, 687, 707, and 945. Contigs were assembled using TIGR's CAP program and a representative EST from each contig was selected for the Unigene set. All singlets were

**ORIGIN**  
 Query Match 3.8%; Score 19; DB 12; Length 469;  
 Best Local Similarity 100.0%; Pred. No. 9.5e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 271 GCCGGCGTGGCGACCACT 289  
 |||||  
**Db** 391 GCCGGCGTGGCGACCACT 409

**RESULT 16**  
**LOCUS** CF608321/c 474 bp mRNA linear EST 30-SEP-2003  
**DEFINITION** GENMA01\_001873 Grape Bud pSPORT1 Library Vitis vinifera cDNA 5', mRNA sequence.

**ACCESSION** CF608321  
**VERSION** CF608321.1 GI:37188972  
**KEYWORDS** EST.  
**SOURCE** Vitis vinifera  
**ORGANISM** Vitis vinifera  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; Vitaceae; Vitis.

**REFERENCE** 1 (bases 1 to 474)  
**AUTHORS** Moser, C., Segala, C., Fontana, P., Salakhudinov, I., Gatto, P., Pindo, M., Zyprian, E., Toepfer, R., Grando, M.S. and Velasco, R.  
**TITLE** Expressed sequence tags from different organs of Vitis vinifera  
**JOURNAL** Unpublished (2003)  
**COMMENT** Contact: Moser C  
 Laboratorio di Genetica Molecolare  
 Istituto Agrario di San Michele all'Adige (IASMA)  
 via E. Mach 1, San Michele all'Adige (TN), I-38010, Italia  
 Tel: 0039-0461-615314  
 Fax: 0039-0461-650956  
 Email: claudio.moser@ismaa.it

The sequencing work has been funded by the 'Fondazione Cassa di Risparmio di Trento e Rovereto',  
 High quality sequence stop: 474.  
**FEATURES** source  
 1. .474  
 Location/Qualifiers  
 /organism="Vitis vinifera"  
 /mol\_type="mRNA"  
 /cultivar="pinot noir"  
 /db\_xref="taxon:29760"  
 /sex="Hermaphrodite"  
 /dev\_stage="bud swelling"  
 /lab\_host="DH10B"  
 /clone\_lib="Grape Bud pSPORT1 Library"  
 /note="Organ: bud; Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI"

**ORIGIN**  
 Query Match 3.8%; Score 19; DB 14; Length 474;  
 Best Local Similarity 100.0%; Pred. No. 9.6e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 221 CCGCCGACCGCGTGATGCG 239  
 |||||  
**Db** 147 CCGCCGACCGCGTGATGCG 129

**RESULT 17**  
**LOCUS** CF608208/c 476 bp mRNA linear EST 30-SEP-2003  
**DEFINITION** GENMA01\_001751 Grape Bud pSPORT1 Library Vitis vinifera cDNA 5', mRNA sequence.

**ACCESSION** CF608208  
**VERSION** CF608208.1 GI:37188859  
**KEYWORDS** EST.  
**SOURCE** Vitis vinifera  
**ORGANISM** Vitis vinifera

Thu Jun 17 09:02:23 2004

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.

1 (bases 1 to 476)  
Moser,C., Segala,C., Fontana,P., Salakhutdinov,I., Gatto,P., Pindo,M., Zyprian,E., Toepfer,R., Grando,M.S. and Velasco,R.  
Expressed sequence tags from different organs of Vitis vinifera  
Unpublished (2003)

CONTACT: Moser C  
Laboratorio di Genetica Molecolare  
Istituto Agrario di San Michele all'Adige (IRASMA)  
Via E. Mach 1, San Michele all'Adige (TN), I-38010, Italia  
Tel: 0039-0461-615314  
Fax: 0039-0461-6150956

Email: claudio.moser@ismaa.it  
The sequencing work has been funded by the 'Fondazione Cassa di Risparmio di Trento e Rovereto'  
High quality sequence stop: 476.

Location/Qualifiers

1..476  
/organism="Vitis vinifera"  
/mol\_type="mRNA"  
/cultivar="Pinot noir"  
/db\_xref="taxon:29760"  
/sex="Hermaphrodite"  
/dev\_stage="bud swelling"  
/lab\_host="DH10B"  
/clone\_lib="Grape Bud pSPORT1 Library"  
/notes="Organ: bud; Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI"

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 476;  
Best Local Similarity 100.0%; Pred. No. 9.6e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCGGACGCGTGTGCG 239

DB 145 CCGCGGACGCGTGTGCG 127

RESULT 18  
BG266809 485 bp mRNA linear EST 20-FEB-2001  
LOCUS 1000106A07.x1 1000 - Unigene I from Maize Genome Project Zea mays  
DEFINITION cDNA, mRNA sequence.

ACCESSION BG266809  
VERSION BG266809.1 GI:12970090  
KEYWORDS EST.

SOURCE Zea mays

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 485)

Walbot.V.  
Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 1000106 row: A column: 07.

Location/Qualifiers

1..485  
/organism="Zea mays"  
/mol\_type="mRNA"  
/db\_xref="dbEST:605025C02.x1"  
/db\_xref="taxon:4577"

## FEATURES

source

Query Match 3.8%; Score 19; DB 14; Length 507;  
Best Local Similarity 100.0%; Pred. No. 9.6e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 GCCGCGACGCGCGCGGTC 489

DB 328 GCCGCGACGCGCGCGGTC 346

## ORIGIN

Query Match 3.8%; Score 19; DB 12; Length 485;

Best Local Similarity 100.0%; Pred. No. 9.6e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 GCCGCGACGCGCGCGGTC 489

DB 328 GCCGCGACGCGCGCGGTC 346

/clone\_lib="1000 - Unigene I from Maize Genome Project"  
/note="This library represents the unique ESTs found in the first round of EST sequencing at Stanford University for the maize genome project. Sequences are present from libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683, 687, 707, and 945. Contigs were assembled using TIGR's CAP program and a representative EST from each contig was selected for the Unigene set. All singlets were also selected."

## ORIGIN

Query Match 3.8%; Score 19; DB 12; Length 485;

Best Local Similarity 100.0%; Pred. No. 9.6e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGCGGTGCCGACCACT 289

DB 412 GCCGCGGTGCCGACCACT 430

## RESULT 19

CB863203

LOCUS HH04A07w HH Hordeum vulgare cDNA clone HH04A07 5-PRIME, mRNA

DEFINITION sequence.

ACCESSION CB863203

VERSION CB863203.1 GI:30057762

KEYWORDS EST.

SOURCE Hordeum vulgare

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 507)

AUTHORS Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and Graner,A.

TITLE Barley ESTs from coleoptile tissue

JOURNAL Unpublished (2003)

COMMENT Contact: Stein Nils

Institute of Plant Genetics and Crop Plant Research (IPK)

Corrensstr. 3, 06466, Gatersleben, Germany

Tel: 039482-5522

Fax: 039482-5595

Email: stein@ipk-gatersleben.de

Insert length: 507

Std Error: 0.00

Plate: 4 row: A column: 7

Seq primer: T7.

Location/Qualifiers

1..507

/organism="Hordeum vulgare"

/mol\_type="mRNA"

/cultivar="Sloop"

/db\_xref="GABI:554860"

/db\_xref="taxon:4513"

/clone="HH04A07"

/tissue\_type="coleoptile"

/dev\_stage="coleoptile, 1 day old"

/lab\_host="DH10B"

/clone\_lib="HH"

/note="Vector: pSPORT; Site 1: SalI (5-end of cDNA); Site 2: NotI (3-end of cDNA); Due to the cloning system used blue/white selection for recombinants is not 100% reliable. Average insert size is 1.3 kb."

## FEATURES

source

Thu Jun 17 09:02:23 2004

us-10-042-665a-3\_copy\_1825\_2325.ol.rst

```

RESULT 20
LOCUS      AI737095      522 bp      mRNA      linear      EST 02-FEB-2000
DEFINITION 605017B11.x2 605 - EndospERM cDNA library from Schmidt lab Zea mays
ACCESSION  AI737095
VERSION     AI737095.1 GI:5058619
KEYWORDS   Zea mays
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 522)
AUTHORS    Walbot V.
TITLE      Zea mays from various cDNA libraries sequenced at Stanford
JOURNAL
COMMENT    Maize ESTs from various cDNA libraries sequenced at Stanford

FEATURES   source
            1..522
            /organism="Zea mays"
            /mol_type="mRNA"
            /cultivar="Ohio43"
            /db_xref="taxon:4577"
            /tissue_type="mucellar, embryo, and endospERM"
            /dev_stage="10-14 days post-pollination"
            /lab_host="DH5(alpha)"
            /clone_lib="605 - EndospERM cDNA library from Schmidt lab"
            /note="Organ: Kernel; Vector: pAD-GAL4-2; Site_1: EcoRI;
            Site_2: XhoI; Kernel endospERM cDNA library from Schmidt
            lab"

ORIGIN
Query Match      3.8%; Score 19; DB 9; Length 522;
Best Local Similarity 100.0%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      271  GCCGGCGTGGCGACCACT 289
            |||||
DB      418  GCCGGCGTGGCGACCACT 436

RESULT 21
LOCUS      AI737056      528 bp      mRNA      linear      EST 02-FEB-2000
DEFINITION 605017B10.x2 605 - EndospERM cDNA library from Schmidt lab Zea mays
ACCESSION  AI737056
VERSION     AI737056.1 GI:5058580
KEYWORDS   Zea mays
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 528)
AUTHORS    Walbot V.
TITLE      Zea mays from various cDNA libraries sequenced at Stanford
JOURNAL
COMMENT    Maize ESTs from various cDNA libraries sequenced at Stanford

FEATURES   source
            1..522
            /organism="Zea mays"
            /mol_type="mRNA"
            /cultivar="Ohio43"
            /db_xref="taxon:4577"
            /tissue_type="mucellar, embryo, and endospERM"
            /dev_stage="10-14 days post-pollination"
            /lab_host="DH5(alpha)"
            /clone_lib="605 - EndospERM cDNA library from Schmidt lab"
            /note="Organ: Kernel; Vector: pAD-GAL4-2; Site_1: EcoRI;
            Site_2: XhoI; Kernel endospERM cDNA library from Schmidt
            lab"

ORIGIN
Query Match      3.8%; Score 19; DB 9; Length 522;
Best Local Similarity 100.0%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      271  GCCGGCGTGGCGACCACT 289
            |||||
DB      418  GCCGGCGTGGCGACCACT 436

RESULT 22
LOCUS      CA401348      530 bp      mRNA      linear      EST 07-NOV-2002
DEFINITION EL01N0419A06.g EndospERM_4 Zea mays cDNA, mRNA sequence.
ACCESSION  CA401348
VERSION     CA401348.1 GI:24766193
KEYWORDS   Zea mays
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 530)
AUTHORS    Lai J., Dey N., Kim C.S., Becraft P., Larkins B., Linton E. and
            Messing J.
TITLE      Sequencing of the maize endospERM ESTs
JOURNAL    Unpublished (2002)
COMMENT    Contact: Lai, Jinsheng
            Dr. Joachim Messing's lab
            Wakeman Institute, Rutgers University
            190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
            Tel: 732-445-3801
            Fax: 732-445-5735
            Email: jlai@waksman.rutgers.edu
            Seq primer: T7.
            Location/Qualifiers
            1..530
            /organism="Zea mays"
            /mol_type="mRNA"
            /cultivar="W22"
            /db_xref="taxon:4577"
            /tissue_type="EndospERM of 7-23DAP"
            /clone_lib="EndospERM_4"
            /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
            XhoI"

ORIGIN
Query Match      3.8%; Score 19; DB 14; Length 530;
Best Local Similarity 100.0%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      271  GCCGGCGTGGCGACCACT 289
            |||||
DB      441  GCCGGCGTGGCGACCACT 459

```

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 605017 row: B column: 10.

Location/Qualifiers

1..528

/organism="Zea mays"

/mol\_type="mRNA"

/cultivar="Ohio43"

/db\_xref="taxon:4577"

/tissue\_type="mucellar, embryo, and endospERM"

/dev\_stage="10-14 days post-pollination"

/lab\_host="DH5(alpha)"

/clone\_lib="605 - EndospERM cDNA library from Schmidt lab"

/note="Organ: Kernel; Vector: pAD-GAL4-2; Site\_1: EcoRI;

Site\_2: XhoI; Kernel endospERM cDNA library from Schmidt

lab"

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 528;

Best Local Similarity 100.0%; Pred. No. 9.6e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289

|||||

DB 409 GCCGGCGTGGCGACCACT 427

## RESULT 22

CA401348

LOCUS

DEFINITION

EL01N0419A06.g EndospERM\_4 Zea mays cDNA, mRNA sequence.

ACCESSION

CA401348

VERSION

CA401348.1 GI:24766193

KEYWORDS

EST.

SOURCE

Zea mays

ORGANISM

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 530)

AUTHORS

Lai J., Dey N., Kim C.S., Becraft P., Larkins B., Linton E. and

Messing J.

TITLE

Sequencing of the maize endospERM ESTs

JOURNAL

Unpublished (2002)

COMMENT

Contact: Lai, Jinsheng

Dr. Joachim Messing's lab

Wakeman Institute, Rutgers University

190 Frelinghuysen Rd., Piscataway, NJ 08854, USA

Tel: 732-445-3801

Fax: 732-445-5735

Email: jlai@waksman.rutgers.edu

Seq primer: T7.

Location/Qualifiers

1..530

/organism="Zea mays"

/mol\_type="mRNA"

/cultivar="W22"

/db\_xref="taxon:4577"

/tissue\_type="EndospERM of 7-23DAP"

/clone\_lib="EndospERM\_4"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI"

ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 530;

Best Local Similarity 100.0%; Pred. No. 9.6e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289

|||||

DB 441 GCCGGCGTGGCGACCACT 459

```

RESULT 23
CF486345/c
LOCUS
DEFINITION POL1_37_E02.b1.A002 Pollen Sorghum bicolor cDNA clone EST 08-SEP-2003
ACCESSION POL1_37_E02.A002 3', mRNA sequence.
VERSION
KEYWORDS
SOURCE
ORGANISM Sorghum bicolor (sorghum)
REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 538)
Cordonnier-Pratt M.-M., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C.,
Sun, F., Sullivan, R., Eastman, A., Cannon, R., Kern, B., Morgan, J.,
Lucas, A., Al-Sheikh, A., Jones, V., Adibi, N., Owen, A., Gao, J. and
Pratt, L.H.
EST database from Sorghum: pollen
Unpublished (2003)
Other ESTs: POL1_37_E02.g1.A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
science; plant material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.
Seq primer: Sug3 (CGACCTGCGACTCGAGCACA)
POLYA=No.

FEATURES
source
1..538
Location/Qualifiers
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="POL1_37_E02.A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Pollen"
/note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI;
Site 2: XhoI; The library was prepared from polyA+ RNA
from pollen at the late vacuolated-vacuolated stage of
development. Pollen was harvested from greenhouse-grown
panicles of sorghum line BTx623. Panicles were removed
from the flag leaf prior to emergence, when no detectable
amylase is present in pollen of male-fertile lines. This
stage represents pollen collected from anthers about 8-14
days prior to anthesis. Double-stranded cDNA was cloned
unidirectionally into different DraIII sites of the
pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG,
3-prime DraIII site is CACCATGTG). XhoI excises the cDNA
insert."

ORIGIN
Query Match 3.8%; Score 19; DB 14; Length 538;
Best Local Similarity 100.0%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289
Db 129 GCCGGCGTGGCGACCACT 111

RESULT 24
AI677252
LOCUS
DEFINITION 605052A08.x1.605 - Endosperm cDNA library from Schmidt lab Zea mays
ACCESSION AI677252
VERSION
KEYWORDS
SOURCE
ORGANISM Zea mays
REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 541)
Walbot, V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605052 row: A column: 08.
Location/Qualifiers
1..541
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="pH5(alpha)"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site 1: EcoRI;
Site 2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

ORIGIN
Query Match 3.8%; Score 19; DB 9; Length 541;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289
Db 391 GCCGGCGTGGCGACCACT 409

RESULT 25
CA401459
LOCUS
DEFINITION E01N0421B10.9 Endosperm_4 Zea mays cDNA, mRNA sequence.
ACCESSION CA401459
VERSION
KEYWORDS
SOURCE
ORGANISM Zea mays
REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 543)
Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and
Messing, J.
Sequencing of the maize endosperm ESTs
Unpublished (2002)
Contact: Lai, Jinsheng
Dr. Joachim Messing's lab
Waksman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735

```

Email: jlai@waksman.rutgers.edu

Seq primer: 17

#### FEATURES

##### source

Location/Qualifiers  
1..543  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="W22"  
/db\_xref="taxon:4577"  
/tissue\_type="Endosperm of 7-23DAP"  
/clone\_lib="Endosperm\_4"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

#### ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 543;  
Best Local Similarity 100.0%; Pred. No. 9.7e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289

DB 455 GCCGGCGTCCGACCACT 473

#### RESULT 26

CF488024/c

LOCUS

DEFINITION POL1\_47 G09.b1 A002 Pollen Sorghum bicolor cDNA clone

POL1\_47 G09 A002 3', mRNA sequence.

CF488024

CF488024.1 GI:34516893

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

544 bp mRNA linear EST 08-SEP-2003  
Sorghum bicolor (sorghum)  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 544)  
Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C.,  
Sun, F., Sullivan, R., Eastman, A., Cannon, R., Kern, B., Morgan, J.,  
Lucas, A., Al-Sheikh, A., Jones, V., Adibi, N., Owen, A., Gao, J. and  
Pratt, L.H.

EST database from Sorghum: pollen

Unpublished (2003)

Other ESTs: POL1\_47 G09.g1 A002

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; plant material and RNA prepared at Texas A & M University;  
sequencing done in the Laboratory for Genomics and Bioinformatics,  
University of Georgia. Sequence ends have been trimmed to exclude  
vector and regions below Phred quality 16. Three-prime sequences  
are presented as their reverse complement and have been trimmed to  
exclude polyA.

Seq primer: Sug3-14 (TAGTCTAGCGCGCGGACC)

POLYA=Yes.

#### FEATURES

##### source

Location/Qualifiers  
1..544  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/cultivar="BTx623"  
/db\_xref="taxon:4558"  
/clone="POL1\_47 G09 A002"  
/lab\_host="DH10B-T1 phage-resistant E. coli"  
/clone\_lib="Pollen"

/note="Organ: Pollen; Vector: pME18S-FL3; Site\_1: XhoI;  
Site\_2: XhoI; The library was prepared from polyA+ RNA  
from pollen at the late vacuolated-vacuolated stage of

#### ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 544;  
Best Local Similarity 100.0%; Pred. No. 9.7e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289

DB 129 GCCGGCGTCCGACCACT 111

#### RESULT 27

CB288531/c

LOCUS

DEFINITION

CB288531

CB288531.1 GI:28602272

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

547 bp mRNA linear EST 27-FEB-2003  
V-B-16B03 VAN-Baker-1 Vitis aestivalis cDNA clone V-B-16B03 5',  
mRNA sequence.

CB288531

CB288531.1 GI:28602272

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 547)  
Hou, H.S., Phamkhanh, T.V., Kovacs, L. and Qiu, W.P.  
Expressed sequence tags of young leaf tissues of a  
disease-resistant *Vitis aestivalis* var. Norton  
Unpublished (2003)  
Contact: Wenping Qiu  
Department of Fruit Science  
Southwest Missouri State University-Mountain Grove  
9740 Red Spring Road, Mountain Grove, MO 65711-2599, USA  
Tel: 417 926 4105  
Fax: 417 926 6646  
Email: weq070f@msu.edu  
Insert Length: 547 Std Error: 0.00  
Plate: VAN-Baker-1-6 row: B column: 03  
Seq primer: T3 PRIMER  
High quality sequence stop: 547  
POLYA=No.

Location/Qualifiers  
1..547  
/organism="Vitis aestivalis"  
/mol\_type="mRNA"  
/cultivar="Norton"  
/db\_xref="taxon:3605"  
/clone="V-B-16B03"  
/tissue\_type="Leaf"  
/dev\_stage="Young leaf"  
/lab\_host="XU10-Gold E. coli"  
/clone\_lib="VAN-Baker-1"  
/note="Vector: pBluescript II SK (+) Phagemid; Site\_1: Xho  
I; Site\_2: EcoR I; VAN-Baker-1 is a cDNA library of Norton  
grape young leaves (*Vitis aestivalis* var. Norton). Norton  
grapevines were grown under normal greenhouse conditions.  
The cDNA synthesis and library construction was performed  
according to the instruction manual for pBluescript II XR  
cDNA library construction kit provided by Stratagene."

#### FEATURES

##### source

Location/Qualifiers  
1..547  
/organism="Vitis aestivalis"  
/mol\_type="mRNA"  
/cultivar="Norton"  
/db\_xref="taxon:3605"  
/clone="V-B-16B03"  
/tissue\_type="Leaf"  
/dev\_stage="Young leaf"  
/lab\_host="XU10-Gold E. coli"  
/clone\_lib="VAN-Baker-1"  
/note="Vector: pBluescript II SK (+) Phagemid; Site\_1: Xho  
I; Site\_2: EcoR I; VAN-Baker-1 is a cDNA library of Norton  
grape young leaves (*Vitis aestivalis* var. Norton). Norton  
grapevines were grown under normal greenhouse conditions.  
The cDNA synthesis and library construction was performed  
according to the instruction manual for pBluescript II XR  
cDNA library construction kit provided by Stratagene."

#### ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 547;  
Best Local Similarity 100.0%; Pred. No. 9.7e+02;

development. Pollen was harvested from greenhouse-grown  
panicles of sorghum line Btx623. Panicles were removed  
from the flag leaf prior to emergence, when no detectable  
amylase is present in pollen of male-fertile lines. This  
stage represents pollen collected from anthers about 8-14  
days prior to anthesis. Double-stranded cDNA was cloned  
unidirectionally into different DraIII sites of the  
pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG,  
3-prime DraIII site is CACCATGTG). XhoI excises the cDNA  
insert."

```

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCCGACGCGTGATGCG 239
Db 114 CCGCCGACGCGTGATGCG 96

RESULT 28
CB288483/c
LOCUS V-B-15E08 VAN-Baker-1 Vitis aestivalis cDNA clone V-B-15E08 5',
DEFINITION mRNA sequence.
ACCESSION CB288483
VERSION CB288483.1 GI:28602224
KEYWORDS EST.
SOURCE Vitis aestivalis
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 562)
AUTHORS Hou,H.S.; Phanikanth,T.V.; Kovacs,L. and Qiu,W.P.
TITLE Expressed sequence tags of young leaf tissues of a
disease-resistant Vitis aestivalis var. Norton
JOURNAL Unpublished (2003)
COMMENT Contact: Wenping Qiu
Department of Fruit Science
Southwest Missouri State University-Mountain Grove
9740 Red Spring Road, Mountain Grove, MO 65711-2999, USA
Tel: 417 926 4105
Fax: 417 926 6646
Email: weq070f@msu.edu
Insert Length: 562 Std Error: 0.00
Plate: VAN-Baker-1-5 row: E column: 08
Seq primer: T3 PRIMER
High quality sequence stop: 562
POLYA=No.

FEATURES
source
location/Qualifiers
1..562
/organism="Vitis aestivalis"
/mol_type="mRNA"
/cultivar="Norton"
/db_xref="taxon:3605"
/clone="V-B-15E08"
/tissue_type="leaf"
/dev_stage="young leaf"
/lab_host="XL10-Gold E.coli"
/clone_lib="VAN-Baker-1"
/note="Vector: pBluescript II SK (+) Phagemid; Site 1: Xho
I; Site 2: EcoR I; VAN-Baker-1 is a cDNA library of Norton
grape young leaves (Vitis aestivalis var. Norton). Norton
grapevines were grown under normal greenhouse conditions.
The cDNA synthesis and library construction was performed
according to the instruction manual for pBluescript II XR
cDNA library construction kit provided by Stratagene."

ORIGIN
Query Match 3.8%; Score 19; DB 14; Length 562;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCCGACGCGTGATGCG 239
Db 142 CCGCCGACGCGTGATGCG 124

RESULT 29
CA398780
LOCUS E01N0309H09.g EndospERM_3 Zea mays cDNA, mRNA sequence.
DEFINITION CA398780
ACCESSION CA398780
VERSION CA398780.1 GI:24763608
KEYWORDS EST.

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCCGACGCGTGATGCG 239
Db 142 CCGCCGACGCGTGATGCG 124

RESULT 30
AI714654
LOCUS 605068G06.xl 605 - EndospERM cDNA library from Schmidt lab Zea mays
DEFINITION cDNA, mRNA sequence.
ACCESSION AI714654
VERSION AI714654.1 GI:5018461
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 571)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 8227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605068 row: G column: 06.

FEATURES
source
location/Qualifiers
1..571
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endospERM"
/dev_stage="10-14 days post-pollination"

```

```

SOURCE Zea mays
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 567)
AUTHORS Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and
Messing,J.
TITLE Sequencing of the maize endospERM ESTs
JOURNAL Unpublished (2002)
COMMENT Contact: Lai, jinsheng
Dr. Joachim Messing's lab
Waksman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlai@waksman.rutgers.edu
Seq primer: T7.
location/Qualifiers
1..567
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W22"
/db_xref="taxon:4577"
/tissue_type="EndospERM of 7-23DAP"
/clone_lib="EndospERM_3"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match 3.8%; Score 19; DB 14; Length 567;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGCGTGCCGACCACCT 289
Db 504 GCCGCGTGCCGACCACCT 522

RESULT 30
AI714654
LOCUS 605068G06.xl 605 - EndospERM cDNA library from Schmidt lab Zea mays
DEFINITION cDNA, mRNA sequence.
ACCESSION AI714654
VERSION AI714654.1 GI:5018461
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 571)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 8227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605068 row: G column: 06.

FEATURES
source
location/Qualifiers
1..571
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endospERM"
/dev_stage="10-14 days post-pollination"

```





```

LOCUS      CB288474          583 bp      mRNA      linear      EST 27-FEB-2003
DEFINITION V-B-15D08 VAN-Baker-1 Vitis aestivalis cDNA clone V-B-15D08 5',
            mRNA sequence.
ACCESSION  CB288474
VERSION    CB288474.1  GI:28602215
KEYWORDS   EST.
SOURCE     Vitis aestivalis
ORGANISM   Vitis aestivalis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; Vitaceae; Vitis.
REFERENCE  1 (bases 1 to 583)
AUTHORS   Hou,H.S., Phanikanth,T.V., Kovacs,L. and Qiu,W.P.
TITLE     Expressed sequence tags of young leaf tissues of a
            disease-resistant Vitis aestivalis var. Norton
JOURNAL    Unpublished (2003)
COMMENT    Contact: Wenping Qiu
            Department of Fruit Science
            Southwest Missouri State University-Mountain Grove
            9740 Red Spring Road, Mountain Grove, MO 65711-2999, USA
            Tel: 417 926 4105
            Fax: 417 926 6646
            Email: weq070@msu.edu
            Insert length: 583 Std Error: 0.00
            Plate: VAN-Baker-1-5 row: D column: 08
            Seq primer: T3 PRIMER
            High quality sequence stop: 583
            POLYA=No.

FEATURES             Location/Qualifiers
     source          1..583
                     /organism="Vitis aestivalis"
                     /mol_type="mRNA"
                     /cultivar="Norton"
                     /db_xref="taxon:3605"
                     /clone="V-B-15D08"
                     /tissue_type="Leaf"
                     /dev_stage="young leaf"
                     /lab_host="Xl10-Gold E .coli"
                     /clone_lib="VAN-Baker-1"
                     /note="Vector: pBluescript II SK (+) Phagemid; Site 1: Xho
                     I; Site 2: EcoR I; VAN-Baker-1 is a cDNA library of Norton
                     grape young leaves (Vitis. aestivalis var. Norton). Norton
                     grapevines were grown under normal greenhouse conditions.
                     The cDNA synthesis and library construction was performed
                     according to the instruction manual for pBluescript II XR
                     cDNA library construction kit provided by Stratagene."

ORIGIN
Query Match      3.8%; Score 19; DB 14; Length 583;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  221  CCGCGCAGCAGCGATGCG 239
Db      |||||||||||||||||||
        CCGCGCAGCAGCGATGCG 129

RESULT 34
LOCUS      BG266929          584 bp      mRNA      linear      EST 20-FEB-2001
DEFINITION 1000108H08.x2 1000 - Unigene I from Maize Genome Project Zea mays
            cDNA, mRNA sequence.
ACCESSION  BG266929
VERSION    BG266929.1  GI:12970330
KEYWORDS   EST.
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 584)
AUTHORS   Walbot,V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford

LOCUS      BG266929          584 bp      mRNA      linear      EST 20-FEB-2000
DEFINITION 605058F01.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION  AI711641
VERSION    AI711641.1  GI:5005579
KEYWORDS   EST.
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 586)
AUTHORS   Walbot,V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford

QY  271  GCCGGCGTGCCGACCACCT 289
Db      |||||||||||||||||||
        GCCGGCGTGCCGACCACCT 405

RESULT 35
LOCUS      AI711641          586 bp      mRNA      linear      EST 02-FEB-2000
DEFINITION 605058F01.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION  AI711641
VERSION    AI711641.1  GI:5005579
KEYWORDS   EST.
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 586)
AUTHORS   Walbot,V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford

QY  387  GCCGGCGTGCCGACCACCT 405
Db      |||||||||||||||||||
        GCCGGCGTGCCGACCACCT 405

FEATURES             Location/Qualifiers
     source          1..584
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="dbEST:605034D06.x1"
                     /db_xref="taxon:4577"
                     /clone_lib="1000 - Unigene I from Maize Genome Project"
                     /note="This library represents the unique ESTs found in
                     the first round of EST sequencing at Stanford University
                     for the maize genome project. Sequences are present from
                     libraries 486, 487, 496, 603, 605, 606, 614, 618, 660,
                     683, 687, 707, and 945. Contigs were assembled using
                     TIGR's CAP program and a representative EST from each
                     contig was selected for the Unigene set. All singlets were
                     also selected."

ORIGIN
Query Match      3.8%; Score 19; DB 12; Length 584;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  271  GCCGGCGTGCCGACCACCT 289
Db      |||||||||||||||||||
        GCCGGCGTGCCGACCACCT 405

FEATURES             Location/Qualifiers
     source          1..586
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /tissue_type="mucellar, embryo, and endosperm"
                     /dev_stage="10-14 days post-pollination"
                     /lab_host="DH5(alpha)"
                     /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
                     /note="Organ: Kernel; Vector: pAD-GAL4-2; Site 1: EcoRI;

```

```
ORIGIN
Query Match      3.8%; Score 19; DB 9; Length 586;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCCGGCGTGGCGACCACT 289
|||||
Db 411 GCCGGCGTGGCGACCACT 429

RESULT 36
Al657469      590 bp      mRNA      linear      EST 02-FEB-2000
LOCUS      605003F08.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
DEFINITION      cDNA, mRNA sequence.
ACCESSION      Al657469
VERSION      Al657469.1 GI:4753559
KEYWORDS      EST.
SOURCE      Zea mays
ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 590)

REFERENCE
AUTHORS      Walbot,V.
TITLE      Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL
COMMENT      Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605003 row: F column: 08.
Location/Qualifiers
1..590
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site 1: EcoRI;
Site 2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

ORIGIN
Query Match      3.8%; Score 19; DB 9; Length 590;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCCGGCGTGGCGACCACT 289
|||||
Db 393 GCCGGCGTGGCGACCACT 411

RESULT 37
CD444963/c
LOCUS      CD444963      601 bp      mRNA      linear      EST 03-JUN-2003
DEFINITION      EL01N0446A06.b Endosperm_4 Zea mays cDNA, mRNA sequence.
ACCESSION      CD444963
VERSION      CD444963.1 GI:31360606
KEYWORDS      EST.
SOURCE      Zea mays
ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
REFERENCE
AUTHORS      Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and
Messing,J.
TITLE      Sequencing of the maize endosperm ESTs
JOURNAL      Unpublished (2002)
COMMENT      Contact: Lai, Jinsheng
Dr. Joachim Messing's lab
Wakman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlai@wakman.rutgers.edu
Seq primer: T3.
Location/Qualifiers
1..601
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W22"
/db_xref="taxon:4577"
/tissue_type="Endosperm of 7-23DAP"
/clone_lib="Endosperm 4"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match      3.8%; Score 19; DB 14; Length 601;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCCGGCGTGGCGACCACT 289
|||||
Db 510 GCCGGCGTGGCGACCACT 492

RESULT 38
Al759026
LOCUS      605085H05.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
DEFINITION      cDNA, mRNA sequence.
ACCESSION      Al759026
VERSION      Al759026.1 GI:5152861
KEYWORDS      EST.
SOURCE      Zea mays
ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 610)
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605085 row: H column: 05.
Location/Qualifiers
1..610
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site_1: EcoRI;
Site_2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

REFERENCE
AUTHORS      Walbot,V.
TITLE      Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL
COMMENT      Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605085 row: H column: 05.
Location/Qualifiers
1..610
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site_1: EcoRI;
Site_2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

FEATURES
source
```

Site\_2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 610;  
Best Local Similarity 100.0%; Pred. No. 9.7e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289  
|||||  
Db 409 GCCGGCGTCCGACCACT 427

## RESULT 39

CB919676/c 616 bp mRNA linear EST 25-APR-2003  
LOCUS  
DEFINITION VVD054D06\_349985 An expressed sequence tag database for abiotic stressed berries of Vitis vinifera var. Chardonnay Vitis vinifera cDNA clone VVD054D06 5, mRNA sequence.

ACCESSION CB919676 GI:30134338

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Vitis vinifera  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

An expressed sequence tag database for abiotic stressed berries of Vitis vinifera var. Chardonnay  
Unpublished (2002)  
Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu

## PCR Primers

FORWARD: T3 20mer

BACKWARD: T7 21mer (backward)

Plate: 054 row: D column: 06

Seq primer: T3 20mer

High quality sequence stop: 616.

## FEATURES

## source

1. .616  
/organism="Vitis vinifera"  
/mol\_type="mRNA"  
/db\_xref="taxon:29760"  
/clone="VVD054D06"  
/tissue\_type="berries"  
/dev\_stage="mixed; 8, 9, 11, 13, 15, 16 weeks daf"  
/clone\_lib="An expressed sequence tag database for abiotic stressed berries of Vitis vinifera var. Chardonnay"  
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 616;  
Best Local Similarity 100.0%; Pred. No. 9.8e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCCGACGACGTGTCG 239  
|||||

Db 24 CCGCCGACGACGTGTCG 6  
|||||

## RESULT 40

AI745808 617 bp mRNA linear EST 02-FEB-2000  
LOCUS  
DEFINITION 605077D11.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays cDNA, mRNA sequence.

AI745808  
AI745808.1 GI:5124184  
EST.

## KEYWORDS

## SOURCE

## ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1 (bases 1 to 617)  
Maize ESTs from various cDNA libraries sequenced at Stanford University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 605077 row: D column: 11.  
Location/Qualifiers

## FEATURES

## source

1. .617  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/tissue\_type="nucellar, embryo, and endosperm"  
/dev\_stage="10-14 days post-pollination"  
/lab\_host="DH5(alpha)"  
/clone\_lib="605 - Endosperm cDNA library from Schmidt lab"  
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site\_1: EcoRI; Site\_2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 617;  
Best Local Similarity 100.0%; Pred. No. 9.8e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289  
|||||

Db 406 GCCGGCGTCCGACCACT 424  
|||||

## RESULT 41

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

AI833967 617 bp mRNA linear EST 02-FEB-2000  
605097G06.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays cDNA, mRNA sequence.  
AI833967  
AI833967.1 GI:5456277  
EST.  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1 (bases 1 to 617)  
Maize ESTs from various cDNA libraries sequenced at Stanford University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 605097 row: G column: 06.  
Location/Qualifiers

## FEATURES

## source

1. .617

```

/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/dev_stage="10-14 days post-pollination"
/lab_host="DHS(alpha)"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/note="Organ: Kernel; Vector: PAD-GAL4-2'; Site 1: EcoRI; Site 2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

ORIGIN
Query Match 3.8%; Score 19; DB 9; Length 617;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCCGCGTGGCGCACCT 289
Db 414 GCCGCGTGGCGCACCT 432

RESULT 42
LOCUS BM356065 621 bp mRNA linear EST 23-JAN-2002
DEFINITION rx33b05.y1 Globodera rostochiensis J2 pCDNAII Smant v1 Globodera
rostochiensis cDNA 5' similar to TR:045080 O45080 C17H12.9 PROTEIN.
[1]; mRNA sequence.
ACCESSION BM356065
VERSION BM356065.1 GI:18090696
KEYWORDS EST.
SOURCE Globodera rostochiensis
ORGANISM Globodera rostochiensis
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heterodermidae; Heteroderinae; Globodera.
REFERENCE 1 (bases 1 to 621)
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
Wyllie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
Tsgaraisvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter, JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was contributed by Dr. Geert Smant of the Laboratory of
Nematology at Wageningen University, Wageningen,
Netherlands (geert.smant@nema.dpw.wau.nl). DNA Sequencing by:
Washington University Genome Sequencing Center

Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 420.
FEATURES
source
1. .621
/organism="Globodera rostochiensis"
/mol_type="mRNA"
/db_xref="taxon:31243"
/dev_stage="J2"
/lab_host="DH10B"
/clone_lib="Globodera rostochiensis J2 pCDNAII Smant v1"
/note="Vector: pCDNAII (Invitrogen); Site 1: BstXI; Site 2: EcoRI; The library was donated for sequencing by Geert Smant from Wageningen University, Laboratory of Nematology, The Netherlands."

ORIGIN
Query Match 3.8%; Score 19; DB 14; Length 625;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;

Qy 178 GCCACCGTGGTGGGGCCA 196
Db 533 GCCACCGTGGTGGGGCCA 551

RESULT 43
LOCUS CD428301/c 625 bp mRNA linear EST 02-JUN-2003
DEFINITION ETH1_28_E08_g1_A002 Ethylene-treated seedlings Sorghum bicolor cDNA
clone ETH1_28_E08_A002 5', mRNA sequence.
ACCESSION CD428301
VERSION CD428301.1 GI:31334770
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 625)
AUTHORS Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S.,
Klein,R.K., Liang,C., Sun,F., Sullivan,R., Shah,M., Salzman,R.,
Chua Tan,N., Gonzalez,M., Lane,S., Miller,V., Nanda,P.,
Olaseinde,O., Eastman,A. and Pratt,L.H.
An EST database from Sorghum: 1-aminocyclopropane-1-carboxylic acid
(ACC)-treated seedlings
Unpublished (2003)
Other ESTs: ETH1_28_E08.bl_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 563 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
Science; plant material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below Phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.
Seq primer: Sug5 (CTTCTGCTCTAAAGCTGCG).
FEATURES
source
1. .625
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="ETH1_28_E08_A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The
library was prepared from polyA+ RNA from seedlings grown
in hydroponic culture. At 8 days of age, medium was
supplemented with 0.5 mM 1-aminocyclopropane-1-carboxylic
acid (ACC) to induce endogenous ethylene (ETH) production.
Roots and shoots were harvested after 27 and 72 hr and
material from both time points was combined prior to RNA
isolation. Double-stranded cDNA was cloned
unidirectionally into different DraIII sites of the
pME18S-FL3 vector (5-prime DraIII site is CACTGTGG,
3-prime DraIII site is CACCAATGTG). XhoI excises the cDNA
insert."
```

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 CTCGGCGCGAGCTGCGG 345  
|||||  
Db 343 CTCGGCGCGAGCTGCGG 325

RESULT 44  
CD442989/c  
LOCUS CD442989 631 bp mRNA linear EST 03-JUN-2003  
DEFINITION EL01N0420H10.b Endosperm\_4 Zea mays cDNA, mRNA sequence.  
ACCESSION CD442989  
VERSION CD442989.1 GI:31358632  
KEYWORDS EST.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 631)  
AUTHORS Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and  
Messing, J.  
TITLE Sequencing of the maize endosperm ESTs  
JOURNAL Unpublished (2002)  
COMMENT Contact: Lai, Jinsheng  
Dr. Joachim Messing's lab  
Waksman Institute, Rutgers University  
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
Tel: 732-445-3801  
Fax: 732-445-5735  
Email: jlai@waksman.rutgers.edu  
Seq primer: T3  
Location/Qualifiers  
1..631  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="W22"  
/db\_xref="taxon:4577"  
/tissue\_type="Endosperm of 7-23DAP"  
/clone\_lib="Endosperm\_4"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI"

FEATURES  
source  
1..636  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="W22"  
/db\_xref="taxon:4577"  
/tissue\_type="Endosperm of 7-23DAP"  
/clone\_lib="Endosperm\_4"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI"

ORIGIN  
Query Match 3.8%; Score 19; DB 14; Length 631;  
Best Local Similarity 100.0%; Pred. No. 9.8e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289  
|||||  
Db 452 GCCGGCGTGGCGACCACT 434

RESULT 45  
CA400622  
LOCUS CA400622 636 bp mRNA linear EST 07-NOV-2002  
DEFINITION EL01N0407F10.9 Endosperm\_4 Zea mays cDNA, mRNA sequence.  
ACCESSION CA400622  
VERSION CA400622.1 GI:24765463  
KEYWORDS EST.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 636)  
AUTHORS Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and  
Messing, J.  
TITLE Sequencing of the maize endosperm ESTs  
JOURNAL Unpublished (2002)  
COMMENT Contact: Lai, Jinsheng  
Dr. Joachim Messing's lab  
Waksman Institute, Rutgers University

190 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
Tel: 732-445-3801  
Fax: 732-445-5735  
Email: jlai@waksman.rutgers.edu  
Seq primer: T7  
Location/Qualifiers  
1..636  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="W22"  
/db\_xref="taxon:4577"  
/tissue\_type="Endosperm of 7-23DAP"  
/clone\_lib="Endosperm\_4"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI"

ORIGIN  
Query Match 3.8%; Score 19; DB 14; Length 636;  
Best Local Similarity 100.0%; Pred. No. 9.8e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289  
|||||  
Db 438 GCCGGCGTGGCGACCACT 456

Search completed: June 16, 2004, 19:29:14  
Job time : 2107.33 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:56:30 ; Search time 241 Seconds  
(without alignments)  
9488.873 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325  
Perfect score: 501  
Sequence: 1 atgttctacagtcgggcac.....cccgctgagggctcgtgc 501

Scoring table: OLIGO.NUC  
Gapop 50.0 , Gapext 50.0

Searched: 2998549 seqs, 2282253817 residues

Word size : 15

Total number of hits satisfying chosen parameters: 2222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PTC\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PTCUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	5.6	692	17	US-10-461-194-11
2	28	5.6	85692	17	US-10-461-194-1
3	26	5.2	88421	9	US-09-976-059-1
4	22	4.4	22	13	US-09-798-033-6
5	22	4.4	22	14	US-10-087-451-6
6	22	4.4	18876	15	US-10-329-079-42
7	22	4.4	61944	15	US-10-329-079-34
8	20	4.0	690	15	US-10-156-761-4780
9	20	4.0	1875	15	US-10-156-761-3143
10	20	4.0	1881	16	US-10-369-493-28460
11	20	4.0	1881	16	US-10-369-493-31220
12	20	4.0	64492	13	US-10-378-083-1
13	20	4.0	9025608	15	US-10-156-761-1
14	20	4.0	9025608	15	US-10-156-761-1

c 15	19	3.8	657	13	US-10-425-114-1800	Sequence 1800, Ap
c 16	19	3.8	1172	9	US-09-924-256A-17	Sequence 17, Appl
c 17	19	3.8	1178	9	US-09-924-256A-91	Sequence 91, Appl
c 18	19	3.8	1212	16	US-10-369-493-40637	Sequence 40637, A
c 19	19	3.8	1818	15	US-10-228-063-7	Sequence 7, Appl
c 20	19	3.8	1915	13	US-10-336-753-46	Sequence 46, Appl
c 21	19	3.8	2223	15	US-10-228-063-9	Sequence 9, Appl
c 22	19	3.8	2263	15	US-10-272-291-1	Sequence 1, Appl
c 23	19	3.8	2263	15	US-10-272-291-2	Sequence 2, Appl
c 24	19	3.8	2267	10	US-09-961-077-25	Sequence 25, Appl
c 25	19	3.8	3738	13	US-10-282-122A-13589	Sequence 13589, A
c 26	19	3.8	4800	15	US-10-272-231-5	Sequence 5, Appl
c 27	19	3.8	4800	17	US-10-109-048-1141	Sequence 1141, Ap
c 28	19	3.8	11100	15	US-10-329-079-5	Sequence 5, Appl
c 29	19	3.8	13315	15	US-10-329-079-1	Sequence 1, Appl
c 30	19	3.8	63158	15	US-10-292-198-1	Sequence 1, Appl
c 31	18	3.6	119	17	US-10-616-390-32	Sequence 32, Appl
c 32	18	3.6	128	17	US-10-616-390-35	Sequence 35, Appl
c 33	18	3.6	175	9	US-09-922-217-873	Sequence 873, App
c 34	18	3.6	175	9	US-09-833-263-873	Sequence 873, App
c 35	18	3.6	175	14	US-10-025-380-873	Sequence 873, App
c 36	18	3.6	243	17	US-10-616-390-4	Sequence 4, Appl
c 37	18	3.6	243	17	US-10-616-390-5	Sequence 5, Appl
c 38	18	3.6	389	17	US-10-616-390-45	Sequence 45, Appl
c 39	18	3.6	418	10	US-09-918-995-7045	Sequence 7045, Ap
c 40	18	3.6	489	13	US-10-342-887-1619	Sequence 1619, Ap
c 41	18	3.6	489	13	US-10-172-118-1619	Sequence 1619, Ap
c 42	18	3.6	489	16	US-10-439-703-36	Sequence 36, Appl
c 43	18	3.6	531	15	US-10-156-761-2447	Sequence 2447, Ap
c 44	18	3.6	613	17	US-10-616-390-6	Sequence 6, Appl
c 45	18	3.6	613	17	US-10-616-390-6	Sequence 6, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-461-194-11  
; Sequence 11, Application US/10461194  
; Publication No. US20040077058A1  
; GENERAL INFORMATION:  
; APPLICANT: Hutchinson, Richard C.  
; APPLICANT: Reid, Ralph C.  
; APPLICANT: Hu, Zhihao  
; APPLICANT: Rascher, Andreas  
; APPLICANT: Schirmer, Andreas  
; APPLICANT: McDaniel, Robert  
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
; TITLE OF INVENTION: PRO-GELEDAMYCIN PRODUCING POLYKETIDE SYNTHASES AND  
; FILE REFERENCE: 300622009700  
; CURRENT APPLICATION NUMBER: US/10/461,194  
; PRIOR FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389,255  
; PRIOR FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/393,929  
; PRIOR FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/395,275  
; PRIOR FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: US 10/212,962  
; PRIOR FILING DATE: 2002-08-05  
; PRIOR APPLICATION NUMBER: US 60/415,326  
; PRIOR FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: US 60/420,820  
; PRIOR FILING DATE: 2002-10-24  
; PRIOR APPLICATION NUMBER: US 60/433,130  
; PRIOR FILING DATE: 2002-12-13  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 692  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus

## US-10-461-194-11

Query Match 5.6%; Score 28; DB 17; Length 692;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 CGCTGTTCCACAGCCTTTCGCACATCG 163  
|||||  
Db 151 CGCTGTTCCACAGCCTTTCGCACATCG 178

## RESULT 2

US-10-461-194-1  
; Sequence 1, Application US/10461194  
; Publication No. US20040077058A1  
; GENERAL INFORMATION:  
; APPLICANT: Hutchinson, Richard C.  
; APPLICANT: Reid, Ralph C.  
; APPLICANT: Hu, Zhihao  
; APPLICANT: Rascher, Andreas  
; APPLICANT: Schirmer, Andreas  
; APPLICANT: McDaniel, Robert  
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
; TITLE OF INVENTION: PRO-CELLULAMICIN PRODUCING POLYPEPTIDE SYNTHASES AND  
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF  
; FILE REFERENCE: 300622009700  
; CURRENT APPLICATION NUMBER: US/10/461,194  
; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389,255  
; PRIOR FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/393,929  
; PRIOR FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/395,275  
; PRIOR FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: US 10/212,962  
; PRIOR FILING DATE: 2002-08-05  
; PRIOR APPLICATION NUMBER: US 60/415,326  
; PRIOR FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: US 60/420,820  
; PRIOR FILING DATE: 2002-10-24  
; PRIOR APPLICATION NUMBER: US 60/433,130  
; PRIOR FILING DATE: 2002-12-13  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 85692  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus  
US-10-461-194-1

Query Match 5.6%; Score 28; DB 17; Length 85692;  
Best Local Similarity 100.0%; Pred. No. 9.8e-05;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 CGCTGTTCCACAGCCTTTCGCACATCG 163  
|||||  
Db 23581 CGCTGTTCCACAGCCTTTCGCACATCG 23608

## RESULT 3

US-09-976-059-1  
; Sequence 1, Application US/09976059  
; Patent No. US2002016474A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Zazopoulos, Emmanuel  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin  
; FILE REFERENCE: 3019-PCT  
; CURRENT APPLICATION NUMBER: US/09/976,059  
; CURRENT FILING DATE: 2001-10-15  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1  
; LENGTH: 88421  
; TYPE: DNA  
; ORGANISM: Actinoplanes sp.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2077)..(3078)  
; OTHER INFORMATION: ORF 1; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (3118)..(4032)  
; OTHER INFORMATION: ORF 2; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (4038)..(5048)  
; OTHER INFORMATION: ORF 3; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (4038)..(5048)  
; OTHER INFORMATION: ORF 3; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (6665)..(5814)  
; OTHER INFORMATION: ORF 4; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (7703)..(6693)  
; OTHER INFORMATION: ORF 5; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (9464)..(8130)  
; OTHER INFORMATION: ORF 6; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (9691)..(10761)  
; OTHER INFORMATION: ORF 7; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (12751)..(10829)  
; OTHER INFORMATION: ORF 8; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (13617)..(12802)  
; OTHER INFORMATION: ORF 9; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (15203)..(13614)  
; OTHER INFORMATION: ORF 10; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (15591)..(15863)  
; OTHER INFORMATION: ORF 11; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (15880)..(19035)  
; OTHER INFORMATION: ORF 12; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (19032)..(39713)  
; OTHER INFORMATION: ORF 13; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (39713)..(65800)  
; OTHER INFORMATION: ORF 14; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (65826)..(66530)  
; OTHER INFORMATION: ORF 15; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (66546)..(67370)  
; OTHER INFORMATION: ORF 16; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (67384)..(70059)  
; OTHER INFORMATION: ORF 17; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (70099)..(70662)  
; OTHER INFORMATION: ORF 18; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (70659)..(71906)  
; OTHER INFORMATION: ORF 19; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (73439)..(71964)  
; OTHER INFORMATION: ORF 20; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (74216)..(73563)  
; OTHER INFORMATION: ORF 21; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (75424)..(74213)

```

; OTHER INFORMATION: ORF 22; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (75535)..(76464)
; OTHER INFORMATION: ORF 23; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (78110)..(76449)
; OTHER INFORMATION: ORF 24; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (79864)..(78107)
; OTHER INFORMATION: ORF 25; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (81624)..(79861)
; OTHER INFORMATION: ORF 26; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (81909)..(81682)
; OTHER INFORMATION: ORF 27; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (82346)..(82062)
; OTHER INFORMATION: ORF 28; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (82587)..(84446)
; OTHER INFORMATION: ORF 29; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (84481)..(85548)
; OTHER INFORMATION: ORF 30; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (85556)..(86845)
; OTHER INFORMATION: ORF 31; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (87372)..(86803)
; OTHER INFORMATION: ORF 32; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (87494)..(88420)
; OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-09-976-059-1

```

```

Query Match      5.2%; Score 26; DB 9; Length 88421;
Best Local Similarity 100.0%; Pred. No. 0.001; 0; Mismatches 0; Indels 0; Gaps 0;
Matches 26; Conservative 0;

```

```

QY 20 CGACCGGGCGGCCCAAGGCGGTGTG 45
DB 38404 CGACCGGGCGGCCCAAGGCGGTGTG 38429

```

```

RESULT 4
US-09-798-033-6
; Sequence 6, Application US/09798033
; Publication No. US20020045220A1
; GENERAL INFORMATION:
; APPLICANT: Koean Biosciences, Inc.
; APPLICANT: Khosla, Chaitan
; APPLICANT: Pfeiffer, Blaine
; TITLE OF INVENTION: BIOSYNTHESIS OF POLYKETIDE SYNTHASE
; FILE REFERENCE: SUBSTRATES
; CURRENT APPLICATION NUMBER: US/09/798,033
; PRIOR FILING DATE: 2003-01-30
; PRIOR APPLICATION NUMBER: 09/687,855
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/159,090
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/206,082
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/232,379
; PRIOR FILING DATE: 2000-09-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

```

```

; OTHER INFORMATION: Primer
US-09-798-033-6
Query Match      4.4%; Score 22; DB 13; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.73; 0; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0;
QY 439 ACCGAGACCTGCGGGCGGATCA 460
DB 1 ACCGAGACCTGCGGGCGGATCA 22
RESULT 5
US-10-087-451-6
; Sequence 6, Application US/10087451
; Publication No. US20020192767A1
; GENERAL INFORMATION:
; APPLICANT: Khosla, Chaitan
; APPLICANT: Pfeiffer, Blaine
; TITLE OF INVENTION: BIOSYNTHESIS OF POLYKETIDE SYNTHASE
; FILE REFERENCE: SUBSTRATES
; CURRENT APPLICATION NUMBER: US/10/087,451
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 09/798,033
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 09/687,855
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/159,090
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/206,082
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/232,379
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/355,211
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
US-10-087-451-6
Query Match      4.4%; Score 22; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.73; 0; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0;
QY 439 ACCGAGACCTGCGGGCGGATCA 460
DB 1 ACCGAGACCTGCGGGCGGATCA 22
RESULT 6
US-10-329-079-42
; Sequence 42, Application US/10329079
; Publication No. US20030198981A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: ZAZOPOULOS, Emmanuel
; APPLICANT: STAFFA, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES
; FILE REFERENCE: 3002-11US
; CURRENT APPLICATION NUMBER: US/10/329,079
; CURRENT FILING DATE: 2002-12-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 42
; LENGTH: 18876
; TYPE: DNA
; ORGANISM: Streptomyces refuineus

```

## US-10-329-079-42

Query Match 4.4%; Score 22; DB 15; Length 18876;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 CGACGGGGCGGCCCAAGGGCGT 41  
Db 13034 CGACGGGGCGGCCCAAGGGCGT 13055

## RESULT 7

US-10-329-079-34  
; Sequence 34, Application US/10329079  
; Publication No. US20030198981A1  
; GENERAL INFORMATION:  
; APPLICANT: FARNET, Chris  
; APPLICANT: ZAZOPOULOS, Emmanuel  
; APPLICANT: STAFFA, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES  
; FILE REFERENCE: 3002-11US  
; CURRENT APPLICATION NUMBER: US/10/329,079  
; CURRENT FILING DATE: 2002-12-24  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 34  
; LENGTH: 61944  
; TYPE: DNA  
; ORGANISM: Streptomyces refuineus  
US-10-329-079-34

Query Match 4.4%; Score 22; DB 15; Length 61944;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 CGACGGGGCGGCCCAAGGGCGT 41  
Db 18094 CGACGGGGCGGCCCAAGGGCGT 18115

## RESULT 8

US-10-156-761-4780  
; Sequence 4780, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 4780  
; LENGTH: 690  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(690)  
US-10-156-761-4780

Query Match 4.0%; Score 20; DB 15; Length 690;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 237 GCGGCTGATCGAGCGCGAGA 256  
Db 261 GCGGCTGATCGAGCGCGAGA 280

## RESULT 9

US-10-156-761-3143  
; Sequence 3143, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 3143  
; LENGTH: 1875  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1875)  
US-10-156-761-3143

Query Match 4.0%; Score 20; DB 15; Length 1875;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 ACCGGGCGGCCCAAGGGCGT 41  
Db 580 ACCGGGCGGCCCAAGGGCGT 599

## RESULT 10

US-10-369-493-28460  
; Sequence 28460, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 28460  
; LENGTH: 1881  
; TYPE: DNA  
; ORGANISM: Burkholderia fungorum  
US-10-369-493-28460

Query Match 4.0%; Score 20; DB 16; Length 1881;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TACAGTCGGGCACGACCGG 26  
|||||  
Db 709 TACAGTCGGGCACGACCGG 728

## RESULT 11

US-10-369-493-31220  
; Sequence 31220, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 31220  
; LENGTH: 1881  
; TYPE: DNA  
; ORGANISM: Burkholderia cepacia  
US-10-369-493-31220

Query Match 4.0%; Score 20; DB 15; Length 1881;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TACAGTCGGGCACGACCGG 26  
|||||  
Db 709 TACAGTCGGGCACGACCGG 728

## RESULT 12

US-10-378-083-1  
; Sequence 1, Application US/10378083  
; Publication No. US20040053274A1  
; GENERAL INFORMATION:  
; APPLICANT: President of Tokyo Institute of Technology  
; TITLE OF INVENTION: A gene cluster of vicenistatin biosynthesis, a vicenistamine  
; TITLE OF INVENTION: glycosyltransferase polypeptide, and a gene encoding the  
; TITLE OF INVENTION: polypeptide  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/378,083  
; CURRENT FILING DATE: 2003-03-04  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 64492  
; TYPE: DNA  
; ORGANISM: Streptomyces halstedii sp.HC-34  
US-10-378-083-1

Query Match 4.0%; Score 20; DB 13; Length 64492;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 ACCGGGGCGGCCCAAGGCGGT 41  
|||||  
Db 42832 ACCGGGGCGGCCCAAGGCGGT 42851

## RESULT 13

US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 4.0%; Score 20; DB 15; Length 9025608;  
Best Local Similarity 100.0%; Pred. No. 0.39;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 GCGGCTGATCGAGCGGCGA 256  
|||||  
Db 5836785 GCGGCTGATCGAGCGGCGA 5836804

## RESULT 14

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 4.0%; Score 20; DB 15; Length 9025608;  
Best Local Similarity 100.0%; Pred. No. 0.39;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 ACCGGGGCGGCCCAAGGCGGT 41  
|||||  
Db 3931605 ACCGGGGCGGCCCAAGGCGGT 3931586

```
RESULT 15
US-10-425-114-1800/c
; Sequence 1800, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53113)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1800
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700193275_FLI
US-10-425-114-1800

Query Match          3.8%; Score 19; DB 13; Length 657;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289
|
Db 239 GCCGGCGTGGCGACCACT 221
|

RESULT 16
US-09-924-256A-17/c
; Sequence 17, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1172
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Clone ps32
US-09-924-256A-17

Query Match          3.8%; Score 19; DB 9; Length 1172;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCTACACGTGGCGACGAC 23
|
Db 1171 TCTACACGTGGCGACGAC 1153
|

RESULT 17
US-09-924-256A-91/c
; Sequence 91, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 91
; LENGTH: 1178
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Clone ps25
US-09-924-256A-91

Query Match          3.8%; Score 19; DB 9; Length 1178;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCTACACGTGGCGACGAC 23
|
Db 1177 TCTACACGTGGCGACGAC 1159
|

RESULT 18
US-10-369-493-40637
; Sequence 40637, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40637
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Caulobacter crescentus
US-10-369-493-40637

Query Match          3.8%; Score 19; DB 16; Length 1212;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 CCTGGCGCGGTGGCGACC 285
|
Db 252 CCTGGCGCGGTGGCGACC 270
|

RESULT 19
US-10-228-063-7/c
; Sequence 7, Application US/10228063
; Publication No. US20030135885A1
; GENERAL INFORMATION:
; APPLICANT: Lanahan, Mike
; TITLE OF INVENTION: Self-processing Plants and Plant Parts
; FILE REFERENCE: 109846.317
```



; CURRENT APPLICATION NUMBER: US/10/228,063  
; CURRENT FILING DATE: 2002-12-12  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 1818  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic  
US-10-228-063-7

Query Match 3.8%; Score 19; DB 15; Length 1818;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACCT 289  
|||  
Db 1655 GCCGGCGTCCGACCACCT 1637

RESULT 20  
US-10-336-753-46/c  
; Sequence 46, Application US/10336753  
; Publication No. US20030226176A1  
; GENERAL INFORMATION:  
; APPLICANT: Guan, Hanning  
; APPLICANT: Keeling, Peter L.  
; TITLE OF INVENTION: PLANT LIKE STARCHES AND THE METHOD OF MAKING THEM IN  
; FILE REFERENCE: 2461-52  
; CURRENT APPLICATION NUMBER: US/10/336,753  
; CURRENT FILING DATE: 2003-01-06  
; PRIOR APPLICATION NUMBER: US/09/402,254  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/06660  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-03  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/042,939  
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-04-04  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 46  
; LENGTH: 1915  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (Join(1..1815, 1819..1914))  
US-10-336-753-46

Query Match 3.8%; Score 19; DB 13; Length 1915;  
Best Local Similarity 100.0%; Pred. No. 8.8;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACCT 289  
|||  
Db 1652 GCCGGCGTCCGACCACCT 1634

RESULT 21  
US-10-228-063-9/c  
; Sequence 9, Application US/10228063  
; Publication No. US20030135885A1  
; GENERAL INFORMATION:  
; APPLICANT: Lahanan, Mike  
; TITLE OF INVENTION: Self-processing Plants and Plant Parts  
; FILE REFERENCE: 109846.317  
; CURRENT APPLICATION NUMBER: US/10/228,063  
; CURRENT FILING DATE: 2002-12-12  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 2223

; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic  
US-10-228-063-9

Query Match 3.8%; Score 19; DB 15; Length 2223;  
Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACCT 289  
|||  
Db 2060 GCCGGCGTCCGACCACCT 2042

RESULT 22  
US-10-272-291-1/c  
; Sequence 1, Application US/10272291  
; Publication No. US20030150023A1  
; GENERAL INFORMATION:  
; APPLICANT: ExSeed Genetics  
; TITLE OF INVENTION: Starch  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/272,291  
; CURRENT FILING DATE: 2002-10-17  
; PRIOR APPLICATION NUMBER: 60/329,525  
; PRIOR FILING DATE: 2001-10-01  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2263  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Wild type sequence EX385  
US-10-272-291-1

Query Match 3.8%; Score 19; DB 15; Length 2263;  
Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACCT 289  
|||  
Db 1845 GCCGGCGTCCGACCACCT 1827

RESULT 23  
US-10-272-291-2/c  
; Sequence 2, Application US/10272291  
; Publication No. US20030150023A1  
; GENERAL INFORMATION:  
; APPLICANT: ExSeed Genetics  
; TITLE OF INVENTION: Starch  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/272,291  
; CURRENT FILING DATE: 2002-10-17  
; PRIOR APPLICATION NUMBER: 60/329,525  
; PRIOR FILING DATE: 2001-10-01  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2263  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: mutation  
; LOCATION: (1643)  
; OTHER INFORMATION: The mutation is at position 1643 in the sequence,  
; OTHER INFORMATION: 1450 bp after the start codon.  
; FEATURE:  
; NAME/KEY: mutation  
; LOCATION: (1643)  
; OTHER INFORMATION: The change is a "G" in the wild type (EX385) to an

US-10-272-291-2/c

; OTHER INFORMATION: "A" in the mutant  
US-10-272-291-2

Query Match 3.8%; Score 19; DB 15; Length 2263;  
Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289  
Db 1845 GCCGGCGTCCGACCACT 1827

RESULT 24

US-09-961-077-25/c  
; Sequence 25, Application US/09961077  
; Publication No. US20030014775A1  
; GENERAL INFORMATION:

APPLICANT: Zwick, Michael G.  
Edington, Brent E.  
McSwiggen, James A.  
Merlo, Patricia Ann Owens  
Guo, Lining  
Skokut, Thomas A.  
Young, Scott A.  
Folkerts, Otto  
Merlo, Donald J.

TITLE OF INVENTION: COMPOSITION AND METHODS FOR  
MODULATION OF GENE EXPRESSION  
IN PLANTS

NUMBER OF SEQUENCES: 1263

CORRESPONDENCE ADDRESS:

ADDRESSEE: LYON & LYON  
STREET: 633 West Fifth Street  
Suite 4700

CITY: Los Angeles

STATE: California

COUNTRY: U.S.A.

ZIP: 90071-2066

COMPUTER READABLE FORM: 3.5" Diskette, 1.44 Mb  
storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: IBM P.C. DOS 5.0

SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA: US/09/961,077

FILING DATE: 21-Sep-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/679,645

FILING DATE: July 12, 1996

APPLICATION NUMBER: 60/001,135

FILING DATE: July 13, 1995

APPLICATION NUMBER: 08/300,726

FILING DATE: September 2, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Waburg, Richard J.

REGISTRATION NUMBER: 32,327

REFERENCE/DOCKET NUMBER: 219/247

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440

TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 2267 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 25:

US-09-961-077-25

Query Match 3.8%; Score 19; DB 10; Length 2267;

Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289  
Db 1840 GCCGGCGTCCGACCACT 1822

RESULT 25

US-10-282-122A-13589  
; Sequence 13589, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:

APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyekind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 13589

LENGTH: 3738

TYPE: DNA

ORGANISM: Burkholderia fungorum

US-10-282-122A-13589

Query Match 3.8%; Score 19; DB 13; Length 3738;

Best Local Similarity 100.0%; Pred. No. 7.6;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 CGGACGGCGCCGCGTCGA 491  
Db 2387 CGGACGGCGCCGCGTCGA 2405

RESULT 26

US-10-272-291-5/c  
; Sequence 5, Application US/10272291  
; Publication No. US20030150023A1  
; GENERAL INFORMATION:

APPLICANT: ExSeed Genetics

; TITLE OF INVENTION: Starch  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/272,291  
; CURRENT FILING DATE: 2002-10-17  
; PRIOR APPLICATION NUMBER: 60/329,525  
; PRIOR FILING DATE: 2001-10-01  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 4800  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: transit\_peptide  
; LOCATION: (1233)..(1448)  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (1554)..(1684)  
; OTHER INFORMATION: number 2  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (1766)..(1859)  
; OTHER INFORMATION: number 3  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (1959)..(2054)  
; OTHER INFORMATION: number 4  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (2145)..(2225)  
; OTHER INFORMATION: number 5  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (2290)..(2412)  
; OTHER INFORMATION: number 6  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (2514)..(2650)  
; OTHER INFORMATION: number 7  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (2761)..(2857)  
; OTHER INFORMATION: number 8  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (3102)..(3211)  
; OTHER INFORMATION: number 9  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (3395)..(3489)  
; OTHER INFORMATION: number 10  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (3682)..(3792)  
; OTHER INFORMATION: number 11  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (3980)..(3976)  
; OTHER INFORMATION: number 12  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: (4106)..(4226)  
; OTHER INFORMATION: number 13  
US-10-272-291-5

Query Match 3.8%; Score 19; DB 15; Length 4800;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACCT 289  
|||  
Db 4056 GCCGGCGTCCGACCACCT 4038

RESULT 27  
US-10-109-048-1141/c  
; Sequence 1141, Application US/10109048  
; Publication No. US20040107461A1  
; GENERAL INFORMATION:  
; APPLICANT: COMMORI, PADMA  
; APPLICANT: KEELING, PETER L.  
; APPLICANT: RAMIREZ, NONA  
; APPLICANT: MCKEAN, ANGELA  
; APPLICANT: GAO, ZHONG  
; APPLICANT: GUAN, HANPING  
; TITLE OF INVENTION: GLUCAN CHAIN LENGTH DOMAINS  
; FILE REFERENCE: 2461-76  
; CURRENT APPLICATION NUMBER: US/10/109,048  
; CURRENT FILING DATE: 2003-03-04  
; PRIOR APPLICATION NUMBER: 60/279,720  
; PRIOR FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 1154  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1141  
; LENGTH: 4800  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-10-109-048-1141

Query Match 3.8%; Score 19; DB 17; Length 4800;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACCT 289  
|||  
Db 4056 GCCGGCGTCCGACCACCT 4038

RESULT 28  
US-10-329-079-5  
; Sequence 5, Application US/10329079  
; Publication No. US20030198981A1  
; GENERAL INFORMATION:  
; APPLICANT: FARNET, Chris  
; APPLICANT: ZAZOFOULOS, Emmanuel  
; APPLICANT: STAFFA, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES  
; FILE REFERENCE: 3002-11US  
; CURRENT APPLICATION NUMBER: US/10/329,079  
; CURRENT FILING DATE: 2002-12-24  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 11100  
; TYPE: DNA  
; ORGANISM: Streptomyces fradiae  
US-10-329-079-5

Query Match 3.8%; Score 19; DB 15; Length 11100;  
Best Local Similarity 100.0%; Pred. No. 5.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 CGACCGGGGGCCCAAGGG 38  
|||  
Db 1958 CGACCGGGGGCCCAAGGG 1976

RESULT 29  
US-10-329-079-1  
; Sequence 1, Application US/10329079  
; Publication No. US20030198981A1  
; GENERAL INFORMATION:  
; APPLICANT: FARNET, Chris  
; APPLICANT: ZAZOFOULOS, Emmanuel  
; APPLICANT: STAFFA, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES

/ FILE REFERENCE: 3002-11US  
/ CURRENT APPLICATION NUMBER: US/10/329,079  
/ CURRENT FILING DATE: 2002-12-24  
/ NUMBER OF SEQ ID NOS: 66  
/ SOFTWARE: PatentIn version 3.0  
/ SEQ ID NO 1  
/ LENGTH: 13315  
/ TYPE: DNA  
/ ORGANISM: Streptomyces fradiae  
US-10-329-079-1

Query Match 3.8%; Score 19; DB 15; Length 13315;  
Best Local Similarity 100.0%; Pred.No. 5.6;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 CGACCGGGCGGCCCAAGGG 38  
DB 4173 CGACCGGGCGGCCCAAGGG 4191

RESULT 30  
US-10-292-198-1/c  
/ Sequence 1, Application US/10292198  
/ Publication No. US20030157654A1  
/ GENERAL INFORMATION:  
/ APPLICANT: SHEN, Ben  
/ APPLICANT: LIU, Wen  
/ TITLE OF INVENTION: BIOSYNTHESIS OF ENEDIYNE COMPOUNDS BY MANIPULATION OF C-1027 GENE  
/ FILE REFERENCE: 054030-0007  
/ CURRENT APPLICATION NUMBER: US/10/292,198  
/ CURRENT FILING DATE: 2003-03-14  
/ PRIOR APPLICATION NUMBER: US 10/159,257  
/ PRIOR FILING DATE: 2002-05-31  
/ PRIOR APPLICATION NUMBER: US 09/478,188  
/ PRIOR FILING DATE: 2000-01-05  
/ PRIOR APPLICATION NUMBER: US 60/115,434  
/ PRIOR FILING DATE: 1999-01-06  
/ NUMBER OF SEQ ID NOS: 146  
/ SOFTWARE: PatentIn version 3.2  
/ SEQ ID NO 1  
/ LENGTH: 63158  
/ TYPE: DNA  
/ ORGANISM: Streptomyces globisporus  
US-10-292-198-1

Query Match 3.8%; Score 19; DB 15; Length 63158;  
Best Local Similarity 100.0%; Pred.No. 3.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 CGACCGGGCGGCCCAAGGG 38  
DB 42110 CGACCGGGCGGCCCAAGGG 42092

RESULT 31  
US-10-616-390-32/c  
/ Sequence 32, Application US/10616390  
/ Publication No. US20040098761A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Trick, Harold N.  
/ APPLICANT: Roe, Judith L.  
/ APPLICANT: Todd, Timothy C.  
/ APPLICANT: Herman, Michael A.  
/ TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes  
/ FILE REFERENCE: KSURF-08151  
/ CURRENT APPLICATION NUMBER: US/10/616,390  
/ CURRENT FILING DATE: 2003-07-09  
/ NUMBER OF SEQ ID NOS: 45  
/ SOFTWARE: PatentIn version 3.2  
/ SEQ ID NO 32  
/ LENGTH: 119  
/ TYPE: DNA

/ ORGANISM: Heterodera glycines  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (41)..(41)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (47)..(47)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (107)..(107)  
/ OTHER INFORMATION: n is a, c, g, or t  
US-10-616-390-32

Query Match 3.6%; Score 18; DB 17; Length 119;  
Best Local Similarity 100.0%; Pred.No. 54;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 CATGAACCGCGCGGACGG 479  
DB 101 CATGAACCGCGCGGACGG 84

RESULT 32  
US-10-616-390-35  
/ Sequence 35, Application US/10616390  
/ Publication No. US20040098761A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Trick, Harold N.  
/ APPLICANT: Roe, Judith L.  
/ APPLICANT: Todd, Timothy C.  
/ APPLICANT: Herman, Michael A.  
/ TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes  
/ FILE REFERENCE: KSURF-08151  
/ CURRENT APPLICATION NUMBER: US/10/616,390  
/ CURRENT FILING DATE: 2003-07-09  
/ NUMBER OF SEQ ID NOS: 45  
/ SOFTWARE: PatentIn version 3.2  
/ SEQ ID NO 35  
/ LENGTH: 128  
/ TYPE: DNA  
/ ORGANISM: Heterodera glycines  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (24)..(24)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (52)..(52)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (84)..(84)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (90)..(90)  
/ OTHER INFORMATION: n is a, c, g, or t  
US-10-616-390-35

Query Match 3.6%; Score 18; DB 17; Length 128;  
Best Local Similarity 100.0%; Pred.No. 53;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 CATGAACCGCGCGGACGG 479  
DB 30 CATGAACCGCGCGGACGG 47

RESULT 33  
US-09-922-217-873  
; Sequence 873, Application US/099222217  
; Patent No. US200200764141  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yuqiu  
; APPLICANT: Smith, Carole E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Skeiky, Yasir A. W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedvick Thomas S.  
; APPLICANT: Carter, Darrick  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; FILE REFERENCE: 210121.471C13  
; CURRENT APPLICATION NUMBER: US/09/922,217  
; CURRENT FILING DATE: 2001-08-03  
; NUMBER OF SEQ ID NOS: 1124  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 873  
; LENGTH: 175  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-922-217-873

Query Match 3.6%; Score 18; DB 9; Length 175;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 CCTGGTGGCGGCGCGCCG 314  
|||||  
Db 36 CCTGGTGGCGGCGCGCCG 53

RESULT 34  
US-09-833-263-873  
; Sequence 873, Application US/09833263  
; Patent No. US20020110547A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Meagher, Madeleine J.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
; FILE REFERENCE: 210121.471C12  
; CURRENT APPLICATION NUMBER: US/09/833,263  
; CURRENT FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 1093  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 873  
; LENGTH: 175  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-833-263-873

Query Match 3.6%; Score 18; DB 9; Length 175;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 CCTGGTGGCGGCGCGCCG 314  
|||||  
Db 36 CCTGGTGGCGGCGCGCCG 53

RESULT 35  
US-10-025-380-873

; Sequence 873, Application US/10025380  
; Publication No. US20020182191A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yuqiu  
; APPLICANT: Smith, Carole E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Skeiky, Yasir A. W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedvick Thomas S.  
; APPLICANT: Carter, Darrick  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; FILE REFERENCE: 210121.471C14  
; CURRENT APPLICATION NUMBER: US/10/025,380  
; CURRENT FILING DATE: 2001-12-19  
; NUMBER OF SEQ ID NOS: 1129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 873  
; LENGTH: 175  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-025-380-873

Query Match 3.6%; Score 18; DB 14; Length 175;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 CCTGGTGGCGGCGCGCCG 314  
|||||  
Db 36 CCTGGTGGCGGCGCGCCG 53

RESULT 36  
US-10-616-390-4/c  
; Sequence 4, Application US/10616390  
; Publication No. US20040098761A1  
; GENERAL INFORMATION:  
; APPLICANT: Trick, Harold N.  
; APPLICANT: Roe, Judith L.  
; APPLICANT: Todd, Timothy C.  
; APPLICANT: Herman, Michael A.  
; TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes  
; FILE REFERENCE: KSURF-08151  
; CURRENT APPLICATION NUMBER: US/10/616,390  
; CURRENT FILING DATE: 2003-07-09  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4  
; LENGTH: 243  
; TYPE: DNA  
; ORGANISM: Heterodera glycines  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (145)..(145)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (151)..(151)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (183)..(183)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:

; NAME/KEY: misc feature  
; LOCATION: (211)..(211)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-616-390-4

Query Match 3.6%; Score 18; DB 17; Length 243;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 CATGAACCGCGCGGACGG 479  
Db 205 CATGAACCGCGCGGACGG 188

## RESULT 37

US-10-616-390-5  
; Sequence 5, Application US/10616390  
; Publication No. US20040098761A1  
; GENERAL INFORMATION:  
; APPLICANT: Trick, Harold N.  
; APPLICANT: Roe, Judith L.  
; APPLICANT: Todd, Timothy C.  
; APPLICANT: Herman, Michael A.  
; TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes  
; FILE REFERENCE: KSURF-08151  
; CURRENT APPLICATION NUMBER: US/10/616,390  
; CURRENT FILING DATE: 2003-07-09  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 5  
; LENGTH: 243  
; TYPE: DNA  
; ORGANISM: Heterodera glycines  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (33)..(33)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (61)..(61)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (93)..(93)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (99)..(99)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-616-390-5

Query Match 3.6%; Score 18; DB 17; Length 243;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 CATGAACCGCGCGGACGG 479  
Db 39 CATGAACCGCGCGGACGG 56

## RESULT 38

US-10-616-390-45/c  
; Sequence 45, Application US/10616390  
; Publication No. US20040098761A1  
; GENERAL INFORMATION:  
; APPLICANT: Trick, Harold N.  
; APPLICANT: Roe, Judith L.  
; APPLICANT: Todd, Timothy C.  
; APPLICANT: Herman, Michael A.  
; TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes  
; FILE REFERENCE: KSURF-08151  
; CURRENT APPLICATION NUMBER: US/10/616,390  
; CURRENT FILING DATE: 2003-07-09

; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 45  
; LENGTH: 389  
; TYPE: DNA  
; ORGANISM: Heterodera glycines  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (248)..(248)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (254)..(254)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (286)..(286)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (314)..(314)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (353)..(383)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-616-390-45

Query Match 3.6%; Score 18; DB 17; Length 389;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 CATGAACCGCGCGGACGG 479  
Db 308 CATGAACCGCGCGGACGG 291

## RESULT 39

US-09-918-995-7045  
; Sequence 7045, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7045  
; LENGTH: 418  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-7045

Query Match 3.6%; Score 18; DB 10; Length 418;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGCGCGCGGCGCGG 367  
Db 309 GCCTGCGCGCGGCGCGG 326

## RESULT 40

US-10-342-887-1619  
; Sequence 1619, Application US/10342887  
; Publication No. US20040058340A1  
; GENERAL INFORMATION:  
; APPLICANT: Dai, Hongyue

```
/ APPLICANT: He, Yudong
/ APPLICANT: Linsley, Peter S.
/ APPLICANT: Mao, Mao
/ APPLICANT: Roberts, Christopher J.
/ APPLICANT: Van 't Veer, Laura Johanna
/ APPLICANT: Van de Vijver, Marc J.
/ APPLICANT: Bernards, Rene
/ TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
/ FILE REFERENCE: 9301-188-999
/ CURRENT APPLICATION NUMBER: US/10/342,887
/ CURRENT FILING DATE: 2003-01-15
/ PRIOR APPLICATION NUMBER: 60/298,918
/ PRIOR FILING DATE: 2001-06-18
/ PRIOR APPLICATION NUMBER: 60/380,710
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 10/172,118
/ PRIOR FILING DATE: 2002-06-14
/ NUMBER OF SEQ ID NOS: 2699
/ SEQ ID NO 1619
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-342-887-1619

Query Match          3.6%; Score 18; DB 13; Length 489;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGGGGGCGCGG 367
DB 321 GCCTGGCGGGGGCGCGG 338

RESULT 41
US-10-172-118-1619
/ Sequence 1619, Application US/10172118
/ Publication No. US20030224374A1
/ GENERAL INFORMATION:
/ APPLICANT: Dai, Hongyue
/ APPLICANT: He, Yudong
/ APPLICANT: Linsley, Peter
/ APPLICANT: Mao, Mao
/ APPLICANT: Roberts, Chris
/ APPLICANT: Van 't Veer, Laura
/ APPLICANT: Van de Vijver, Marc
/ APPLICANT: Bernards, Rene
/ TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
/ FILE REFERENCE: 9301-175-999
/ CURRENT APPLICATION NUMBER: US/10/172,118
/ CURRENT FILING DATE: 2002-06-14
/ PRIOR APPLICATION NUMBER: 60/380,770
/ PRIOR FILING DATE: 2002-05-14
/ NUMBER OF SEQ ID NOS: 2699
/ SEQ ID NO 1619
/ LENGTH: 489
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ PUBLICATION INFORMATION:
/ DATABASE ACCESSION NUMBER: NM 016199
/ DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-1619

Query Match          3.6%; Score 18; DB 13; Length 489;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGGGGGCGCGG 367
DB 321 GCCTGGCGGGGGCGCGG 338

RESULT 42
US-10-439-703-36
```

```
/ Sequence 36, Application US/10439703
/ Publication No. US20040018527A1
/ GENERAL INFORMATION:
/ APPLICANT: Chang, Jenny
/ APPLICANT: O'Connell, Peter
/ TITLE OF INVENTION: Differential Patterns of Gene Expression that Predict for Docetax
/ TITLE OF INVENTION: Chemosensitivity and Chemoresistance
/ FILE REFERENCE: HO-P2482US1/10205813
/ CURRENT APPLICATION NUMBER: US/10/439,703
/ CURRENT FILING DATE: 2003-05-16
/ PRIOR APPLICATION NUMBER: US 60/381,141
/ PRIOR FILING DATE: 2002-05-17
/ NUMBER OF SEQ ID NOS: 91
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 36
/ LENGTH: 489
/ TYPE: DNA
/ ORGANISM: Human
US-10-439-703-36

Query Match          3.6%; Score 18; DB 16; Length 489;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGGGGGCGCGG 367
DB 321 GCCTGGCGGGGGCGCGG 338

RESULT 43
US-10-156-761-2447
/ Sequence 2447, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ CURRENT FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 2447
/ LENGTH: 531
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(531)
US-10-156-761-2447

Query Match          3.6%; Score 18; DB 15; Length 531;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 CGACCTTCCTGGCGCGG 277
DB 101 CGACCTTCCTGGCGCGG 118

RESULT 44
US-10-616-390-6
/ Sequence 6, Application US/10616390
/ Publication No. US20040098761A1
/ GENERAL INFORMATION:
```



```
/ APPLICANT: Trick, Harold N.
/ APPLICANT: Roe, Judith L.
/ APPLICANT: Todd, Timothy C.
/ APPLICANT: Herman, Michael A.
/ TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes
/ FILE REFERENCE: KSURF-08151
/ CURRENT APPLICATION NUMBER: US/10/616,390
/ CURRENT FILING DATE: 2003-07-09
/ NUMBER OF SEQ ID NOS: 45
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 6
/ LENGTH: 613
/ TYPE: DNA
/ ORGANISM: Heterodera glycines
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (45)..(45)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (73)..(73)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (105)..(105)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (111)..(111)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (509)..(509)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (515)..(515)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (547)..(547)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (575)..(575)
/ OTHER INFORMATION: n is a, c, g, or t
/ US-10-616-390-6
/
/ Query Match 3.6%; Score 18; DB 17; Length 613;
/ Best Local Similarity 100.0%; Pred. No. 37;
/ Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 462 CATGAACCCCGCGGACGG 479
/ DB 51 CATGAACCCCGCGGACGG 68
/
/ RESULT 45
/ US-10-616-390-6/c
/ Sequence 6, Application US/10616390
/ Publication No. US20040098761A1
/ GENERAL INFORMATION:
/ APPLICANT: Trick, Harold N.
/ APPLICANT: Roe, Judith L.
/ APPLICANT: Todd, Timothy C.
/ APPLICANT: Herman, Michael A.
/ TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes
/ FILE REFERENCE: KSURF-08151
/ CURRENT APPLICATION NUMBER: US/10/616,390
/ CURRENT FILING DATE: 2003-07-09
/ NUMBER OF SEQ ID NOS: 45
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 6
```

```
/ LENGTH: 613
/ TYPE: DNA
/ ORGANISM: Heterodera glycines
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (45)..(45)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (73)..(73)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (105)..(105)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (111)..(111)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (509)..(509)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (515)..(515)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (547)..(547)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (575)..(575)
/ OTHER INFORMATION: n is a, c, g, or t
/ US-10-616-390-6
/
/ Query Match 3.6%; Score 18; DB 17; Length 613;
/ Best Local Similarity 100.0%; Pred. No. 37;
/ Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 462 CATGAACCCCGCGGACGG 479
/ DB 569 CATGAACCCCGCGGACGG 552
/
/ Search completed: June 16, 2004, 20:03:04
/ Job time : 265 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 16, 2004, 15:48:58 ; Search time 57.3333 Seconds  
(without alignments)  
4849.367 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325  
Perfect score: 501  
Sequence: 1 atgtttacagctcggcac.....ccgcgctcgggctgtgc 501

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 15

Total number of hits satisfying chosen parameters: 521

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	4.0	5275	1 US-08-485-588-1	Sequence 1, Appli
2	20	4.0	5275	1 US-08-484-565-1	Sequence 1, Appli
3	20	4.0	5275	2 US-08-480-751-1	Sequence 1, Appli
4	20	4.0	5275	2 US-08-943-986-1	Sequence 1, Appli
5	20	4.0	5275	3 US-08-353-784-1	Sequence 1, Appli
6	20	4.0	5275	3 US-08-484-719B-1	Sequence 1, Appli
7	20	4.0	5275	4 US-08-484-159-1	Sequence 1, Appli
8	19	3.8	1172	3 US-08-861-774E-17	Sequence 17, Appli
9	19	3.8	1178	3 US-08-861-774E-91	Sequence 91, Appli
10	19	3.8	1818	4 US-09-731-166-3	Sequence 3, Appli
11	19	3.8	2267	4 US-08-679-645-25	Sequence 25, Appli
12	19	3.8	4800	3 US-08-941-443A-4	Sequence 4, Appli
13	18	3.6	495	4 US-09-621-976-1819	Sequence 1819, Ap
14	18	3.6	497	4 US-09-621-976-797	Sequence 797, Appli
15	18	3.6	498	2 US-08-722-349-2	Sequence 2, Appli
16	18	3.6	498	3 US-09-204-328-2	Sequence 2, Appli
17	18	3.6	789	4 US-09-252-991A-6590	Sequence 6590, Ap
18	18	3.6	942	4 US-09-252-991A-6619	Sequence 6619, Ap
19	18	3.6	954	4 US-09-252-991A-8902	Sequence 8902, Ap
20	18	3.6	2469	4 US-09-252-991A-6679	Sequence 6679, Ap
21	18	3.6	3111	4 US-09-252-991A-8790	Sequence 8790, Ap
22	18	3.6	77536	4 US-09-410-551B-1	Sequence 1, Appli
23	18	3.6	4403765	3 US-09-103-840A-2	Sequence 2, Appli
24	18	3.6	4411529	3 US-09-103-840A-1	Sequence 1, Appli
25	17	3.4	333	3 US-09-060-756-318	Sequence 318, App
26	17	3.4	333	4 US-09-670-314-318	Sequence 318, App
27	17	3.4	456	4 US-09-252-991A-9308	Sequence 9308, Ap

ALIGNMENTS

RESULT 1

US-08-485-588-1  
; Sequence 1, Application US/08485588  
; Patent No. 5688938

GENERAL INFORMATION:

APPLICANT: Edward M. Brown  
APPLICANT: Steven C. Hebert  
APPLICANT: Forrest H. Fuller  
APPLICANT: James E. Garrett, Jr.  
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
MOLECULES  
NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon  
STREET: First Interstate World Center  
STREET: Suite 4700  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: FASTSEQ  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,588  
FILING DATE: 7 June, 1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below: 9

APPLICATION NUMBER: 08/353,784

FILING DATE: 9 December, 1994

APPLICATION NUMBER: PCT/US/94/12117

FILING DATE: 21 October, 1994

APPLICATION NUMBER: U.S. 08/292,827

FILING DATE: 23 August, 1994

APPLICATION NUMBER: U.S. 08/141,248

FILING DATE: 22 October, 1993

APPLICATION NUMBER: U.S. 08/009,389

FILING DATE: 23 February, 1993

APPLICATION NUMBER: U.S. 08/017,127

FILING DATE: 12 February, 1993

APPLICATION NUMBER: U.S. 07/934,161

FILING DATE: 21 August, 1992

APPLICATION NUMBER: U.S. 07/834,044

FILING DATE: 11 February, 1992

APPLICATION NUMBER: U.S. 07/749,451

Sequence 680, App  
Sequence 680, App  
Sequence 9256, Ap  
Sequence 4264, Ap  
Sequence 2, Appli  
Sequence 15003, A  
Sequence 6751, Ap  
Sequence 1, Appli  
Sequence 5, Appli  
Sequence 1, Appli  
Sequence 6874, Ap  
Sequence 4787, Ap  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 9291, Ap  
Sequence 9280, Ap  
Sequence 1, Appli  
Sequence 2, Appli

```
/ FILING DATE: 23 August, 1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Heber, Sheldon O.
/ REGISTRATION NUMBER: 38,179
/ REFERENCE/DOCKET NUMBER: 213/005
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5275 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA to mRNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 515..3769
/ OTHER INFORMATION:
/ US-08-485-588-1

Query Match 4.0%; Score 20; DB 1; Length 5275;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGCGCGGG 381
Db 223 GCGCGGTCTCGCGCGGG 242

RESULT 2
US-08-484-565-1
; Sequence 1, Application US/08484565
; Patent No. 5763569
; GENERAL INFORMATION:
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,565
; FILING DATE: 7 June, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 9
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
```

```
/ APPLICATION NUMBER: U.S. 08/017,127
/ FILING DATE: 12 February, 1993
/ APPLICATION NUMBER: U.S. 07/934,161
/ FILING DATE: 21 August, 1992
/ APPLICATION NUMBER: U.S. 07/834,044
/ FILING DATE: 11 February, 1992
/ APPLICATION NUMBER: U.S. 07/749,451
/ FILING DATE: 23 August, 1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Heber, Sheldon O.
/ REGISTRATION NUMBER: 38,179
/ REFERENCE/DOCKET NUMBER: 213/006
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5275 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA to mRNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 515..3769
/ OTHER INFORMATION:
/ US-08-484-565-1

Query Match 4.0%; Score 20; DB 1; Length 5275;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGCGCGGG 381
Db 223 GCGCGGTCTCGCGCGGG 242

RESULT 3
US-08-480-751-1
; Sequence 1, Application US/08480751
; Patent No. 5858684
; GENERAL INFORMATION:
; APPLICANT: Edward F. Nemeth
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: Forrest H. Fuller
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,751
; FILING DATE: 7 June, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 9
; APPLICATION NUMBER: 08/353,784
```

```

; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
; REGISTRATION NUMBER: 38,179
; REFERENCE/DOCKET NUMBER: 213/004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 515..3769
; OTHER INFORMATION:
;
US-08-480-751-1

```

```

Query Match 4.0%; Score 20; DB 2; Length 5275;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 362 GCGGGTCTCGGCGCGGG 381
Db 223 GCGGGTCTCGGCGCGGG 242

```

```

RESULT 4
US-08-943-986-1
; Sequence 1, Application US/08943986
; Patent No. 5962314
; GENERAL INFORMATION:
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/943,986
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; FILING DATE: 7-June-1995
; APPLICATION NUMBER: 08/484,565
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
; REGISTRATION NUMBER: 38,179
; REFERENCE/DOCKET NUMBER: 213/006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 515..3769
; OTHER INFORMATION:
;
US-08-943-986-1

```

```

Query Match 4.0%; Score 20; DB 2; Length 5275;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 362 GCGGGTCTCGGCGCGGG 381
Db 223 GCGGGTCTCGGCGCGGG 242

```

```

RESULT 5
US-08-353-784-1
; Sequence 1, Application US/08353784
; Patent No. 6011068
; GENERAL INFORMATION:
; APPLICANT: Edward F. Nemeth, Edward M.
; APPLICANT: Brown, Steven C. Hebert,
; APPLICANT: Bradford C. Van Wagenen, Manuel
; APPLICANT: F. Balandrin, Forrest H. Fuller,
; APPLICANT: Eric G. DelMar, and Scott T. Moe
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles

```

Thu Jun 17 09:02:22 2004

```

; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,784
; FILING DATE: 9 December, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 8
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
; REGISTRATION NUMBER: 38,179
; REFERENCE/DOCKET NUMBER: 209/069
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 515..3769
; OTHER INFORMATION:
; US-08-353-784-1
;
; Query Match 4.0%; Score 20; DB 3; Length 5275;
; Best Local Similarity 100.0%; Pred. No. 2.5;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 362 GCGGGTCTCGGCGCGG 381
; Db 223 GCGGGTCTCGGCGCGG 242
;
; RESULT 6
; US-08-484-719B-1
; Sequence 1, Application US/08484719B
; Patent No. 6031003
; GENERAL INFORMATION:
; APPLICANT: Edward F. Nemeth, Edward M.
; APPLICANT: Brown, Steven C. Hebert,
; APPLICANT: Bradford C. Van Wagenen,
; APPLICANT: Manuel F. Balandrin,
; APPLICANT: Forrest H. Fuller, Eric G.
; APPLICANT: Delmar, Scott T. Moe
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE

```

```

; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS Word
; SOFTWARE: FASTSEQ for Windows Version 3.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,719B
; FILING DATE: 7 June, 1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Douglas C. Murdock
; REGISTRATION NUMBER: 37,549
; REFERENCE/DOCKET NUMBER: 213/007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 515..3769
; OTHER INFORMATION:
; US-08-484-719B-1
;
; Query Match 4.0%; Score 20; DB 3; Length 5275;
; Best Local Similarity 100.0%; Pred. No. 2.5;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 362 GCGGGTCTCGGCGCGG 381
; Db 223 GCGGGTCTCGGCGCGG 242
;
; RESULT 7
; US-08-484-159-1
; Sequence 1, Application US/08484159
; Patent No. 6313146
; GENERAL INFORMATION:

```

APPLICANT: Bradford C. Van Wagenen  
APPLICANT: Manuel F. Balandrin  
APPLICANT: Eric G. Del Mar  
APPLICANT: Edward F. Nemeth  
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
TITLE OF INVENTION: MOLECULES  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Lyon & Lyon  
STREET: First Interstate World Center  
STREET: Suite 4700  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: FASTSEQ  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,159  
FILING DATE: 7 June, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below: 9  
APPLICATION NUMBER: 08/353,784  
FILING DATE: 9 December, 1994  
APPLICATION NUMBER: PCT/US/94/12117  
FILING DATE: 21 October, 1994  
APPLICATION NUMBER: U.S. 08/292,827  
FILING DATE: 23 August, 1994  
APPLICATION NUMBER: U.S. 08/141,248  
FILING DATE: 22 October, 1993  
APPLICATION NUMBER: U.S. 08/009,389  
FILING DATE: 23 February, 1993  
APPLICATION NUMBER: U.S. 08/017,127  
FILING DATE: 12 February, 1993  
APPLICATION NUMBER: U.S. 07/934,161  
FILING DATE: 21 August, 1992  
APPLICATION NUMBER: U.S. 07/834,044  
FILING DATE: 11 February, 1992  
APPLICATION NUMBER: U.S. 07/749,451  
FILING DATE: 23 August, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Heber, Sheldon O.  
REGISTRATION NUMBER: 38,179  
REFERENCE/DOCKET NUMBER: 214/101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5275 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 515..3769  
OTHER INFORMATION:  
US-08-484-159-1

Query Match 4.0%; Score 20; DB 4; Length 5275;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCTCGGCGCGGG 381  
|||||

Db 223 GCGCGGTCTCTCGGCGCGGG 242

## RESULT 8

US-08-861-774E-17/c  
; Sequence 17, Application US/08861774E  
; Patent No. 6297007  
; GENERAL INFORMATION:  
; APPLICANT: Waters, Barbara  
; APPLICANT: Miao, Vivian  
; APPLICANT: Ho, Yap  
; APPLICANT: Tong, Seow  
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR  
; TITLE OF INVENTION: BIOACTIVE MOLECULES  
; FILE REFERENCE: 9993-006  
; CURRENT APPLICATION NUMBER: US/08/861,774E  
; CURRENT FILING DATE: 1997-05-22  
; NUMBER OF SEQ ID NOS: 94  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 1172  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Clone ps32  
US-08-861-774E-17

Query Match 3.8%; Score 19; DB 3; Length 1172;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCTACACGTCTCGGCGAC 23  
|||||

Db 1171 TCTACACGTCTCGGCGAC 1153

## RESULT 9

US-08-861-774E-91/c  
; Sequence 91, Application US/08861774E  
; Patent No. 6297007  
; GENERAL INFORMATION:  
; APPLICANT: Waters, Barbara  
; APPLICANT: Miao, Vivian  
; APPLICANT: Ho, Yap  
; APPLICANT: Tong, Seow  
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR  
; TITLE OF INVENTION: BIOACTIVE MOLECULES  
; FILE REFERENCE: 9993-006  
; CURRENT APPLICATION NUMBER: US/08/861,774E  
; CURRENT FILING DATE: 1997-05-22  
; NUMBER OF SEQ ID NOS: 94  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 91  
; LENGTH: 1178  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Clone ps25  
US-08-861-774E-91

Query Match 3.8%; Score 19; DB 3; Length 1178;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCTACACGTCTCGGCGAC 23  
|||||

Db 1177 TCTACACGTCTCGGCGAC 1159

## RESULT 10

US-09-731-166-3/c  
; Sequence 3, Application US/09731166  
; Patent No. 6639126

; GENERAL INFORMATION:
; APPLICANT: Sewalt, Vincent J. H.
; APPLICANT: Singletary, George W.
; TITLE OF INVENTION: Production of Modified Polysaccharides
; FILE REFERENCE: 35718/206348
; CURRENT APPLICATION NUMBER: US/09/731,166
; CURRENT FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/169,993
; PRIOR FILING DATE: 1999-12-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1818
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: UDP-glucose starch glycosyl transferase -- Genbank
; OTHER INFORMATION: Accession No. 6639126 22509
; NAME/KEY: CDS
; LOCATION: (1)...(1818)
US-09-731-166-3

Query Match 3.8%; Score 19; DB 4; Length 1818;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCGGGGTGCGGACCACT 289
Db 1652 GCGGGGTGCGGACCACT 1634

RESULT 11
US-08-679-645-25/c
; Sequence 25, Application US/08679645
; Patent No. 6350934
; GENERAL INFORMATION:
; APPLICANT: Zwick, Michael G.
; APPLICANT: Edington, Brent E.
; APPLICANT: McSwiggen, James A.
; APPLICANT: Merlo, Patricia Ann Owens
; APPLICANT: Guo, Lining
; APPLICANT: Skokut, Thomas A.
; APPLICANT: Young, Scott A.
; APPLICANT: Folkerts, Otto
; APPLICANT: Merlo, Donald J.
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR
; TITLE OF INVENTION: MODULATION OF GENE EXPRESSION
; TITLE OF INVENTION: IN PLANTS
; NUMBER OF SEQUENCES: 1263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/679,645
; FILING DATE: July 12, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,135
; FILING DATE: July 13, 1995
; APPLICATION NUMBER: 08/300,726

; FILING DATE: September 2, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 219/247
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2267 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-679-645-25

Query Match 3.8%; Score 19; DB 4; Length 2267;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCGGGGTGCGGACCACT 289
Db 1840 GCGGGGTGCGGACCACT 1822

RESULT 12
US-08-941-445A-4/c
; Sequence 4, Application US/08941445A
; Patent No. 6107060
; GENERAL INFORMATION:
; APPLICANT: Keeling, Peter
; APPLICANT: Guan, Haining
; TITLE OF INVENTION: Starch Encapsulation
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/941,445A
; FILING DATE: 30-SEP-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,855
; FILING DATE: 30-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Winner, Ellen P
; REGISTRATION NUMBER: 28,547
; REFERENCE/DOCKET NUMBER: 89-97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS



LOCATION: join(1449..1553, 1685..1765, 1860..1958, 2055  
LOCATION: ..2144, 2226..2289, 2413..2513, 2651..2760, 2858  
LOCATION: ..3101, 3212..3394, 3490..3681, 3793..3879, 3977  
LOCATION: ..4105, 4227..4343)  
US-08-941-445A-4

Query Match 3.8%; Score 19; DB 3; Length 4800;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACACCT 289  
|||||  
DB 4056 GCCGGCGTGGCGACACCT 4038  
|||||

## RESULT 13

US-09-621-976-1819  
; Sequence 1819, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 1819  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 21..335  
; NAME/KEY: sig.peptide  
; LOCATION: 21..191  
; OTHER INFORMATION: Von Heijne matrix  
; OTHER INFORMATION: score 4.4000009536743  
; OTHER INFORMATION: seq STLCTAPLSTCG/EP  
US-09-621-976-1819

Query Match 3.6%; Score 18; DB 4; Length 495;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGCGGGCGCGG 367  
|||||  
DB 338 GCCTGGCGCGGGCGCGG 355  
|||||

## RESULT 14

US-09-621-976-797  
; Sequence 797, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 797  
; LENGTH: 497  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 182..496

## US-09-621-976-797

Query Match 3.6%; Score 18; DB 4; Length 497;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGCGGGCGCGG 367  
|||||  
DB 330 GCCTGGCGCGGGCGCGG 347  
|||||

## RESULT 15

US-08-722-349-2  
; Sequence 2, Application US/08722349  
; Patent No. 5955299  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga B.  
; APPLICANT: Zweiger, Gary B.  
; TITLE OF INVENTION: NOVEL SNRNP SM PROTEINS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/722,349  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0132 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 498 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; LIBRARY:  
; CLONE: Consensus  
US-08-722-349-2

Query Match 3.6%; Score 18; DB 2; Length 498;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGCGGGCGCGG 367  
|||||  
DB 339 GCCTGGCGCGGGCGCGG 356  
|||||

## RESULT 16

US-09-204-328-2  
; Sequence 2, Application US/09204328  
; Patent No. 6090564  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.

APPLICANT: Bandman, Olga  
APPLICANT: Zweiger, Gary B.  
TITLE OF INVENTION: NOVEL SNRP SM PROTEINS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/204,328  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,349  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0132 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 498 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
LIBRARY:  
CLONE: Consensus  
US-09-204-328-2

Query Match 3.6%; Score 18; DB 3; Length 498;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 350 GCCTGGCCGGGGCGCG 367  
Db 339 GCCTGGCCGGGGCGCG 356

RESULT 17  
US-09-252-991A-6590  
; Sequence 6590, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT FILING DATE: 1999-02-18  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6590  
; LENGTH: 789  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6590  
Query Match 3.6%; Score 18; DB 4; Length 789;

Best Local Similarity 100.0%; Pred. No. 27;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 266 TCCTGGCCGGGTGCCGA 283  
Db 159 TCCTGGCCGGGTGCCGA 176

RESULT 18  
US-09-252-991A-6619  
; Sequence 6619, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT FILING DATE: 1999-02-18  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6619  
; LENGTH: 942  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6619

Query Match 3.6%; Score 18; DB 4; Length 942;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 266 TCCTGGCCGGGTGCCGA 283  
Db 776 TCCTGGCCGGGTGCCGA 793

RESULT 19  
US-09-252-991A-8902  
; Sequence 8902, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT FILING DATE: 1999-02-18  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8902  
; LENGTH: 954  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8902

Query Match 3.6%; Score 18; DB 4; Length 954;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 342 GCGGATCGGCTGGCCGG 359  
Db 735 GCGGATCGGCTGGCCGG 752

RESULT 20  
US-09-252-991A-6679/c  
; Sequence 6679, Application US/09252991A

```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6679
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6679

Query Match          3.6%; Score 18; DB 4; Length 2469;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 TCCTGGCGCGGTGCCGA 283
   |||||
Db 2045 TCCTGGCGCGGTGCCGA 2028

RESULT 21
US-09-252-991A-8790
; Sequence 8790, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8790
; LENGTH: 3111
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8790

Query Match          3.6%; Score 18; DB 4; Length 3111;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 GCGGATCGGCTGGCCGG 359
   |||||
Db 2146 GCGGATCGGCTGGCCGG 2163

RESULT 22
US-09-410-551B-1/c
; Sequence 1, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
```

```
; CURRENT APPLICATION NUMBER: US/09/410.551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275)...(71465)
US-09-410-551B-1

Query Match          3.6%; Score 18; DB 4; Length 77536;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 CCCGCTGTCCACAGCCT 152
   |||||
Db 45947 CCCGCTGTCCACAGCCT 45930

RESULT 23
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          3.6%; Score 18; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 CGCGGACGGCGCGCGG 487
   |||||
Db 719415 CGCGGACGGCGCGCGG 719398

RESULT 24
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
```

;; TITLE OF INVENTION: TUBERCULOSIS  
;; FILE REFERENCE: 24366-20007.00  
;; CURRENT APPLICATION NUMBER: US/09/103,840A  
;; CURRENT FILING DATE: 1998-06-24  
;; NUMBER OF SEQ ID NOS: 2  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 1  
;; LENGTH: 4411529  
;; TYPE: DNA  
;; ORGANISM: Mycobacterium tuberculosis  
;; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 3.6%; Score 18; DB 3; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 CGCGGACGGCGCGCG 487  
DB 717982 CGCGGACGGCGCGCG 717965

RESULT 25  
US-09-060-756-318/c  
; Sequence 318, Application US/09060756  
; Patent No. 6183957  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/060,756  
; CURRENT FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 318  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-09-060-756-318

Query Match 3.4%; Score 17; DB 3; Length 333;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 GACCAGGACGGGTGCT 125  
DB 71 GACCAGGACGGGTGCT 55

RESULT 26  
US-09-670-314-318/c  
; Sequence 318, Application US/09670314  
; Patent No. 6492506  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/670,314  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 09/060,756  
; PRIOR FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: PatentIn Ver. 2.0

;; SEQ ID NO 318  
;; LENGTH: 333  
;; TYPE: DNA  
;; ORGANISM: Mycobacterium tuberculosis  
US-09-670-314-318

Query Match 3.4%; Score 17; DB 4; Length 333;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 GACCAGGACGGGTGCT 125  
DB 71 GACCAGGACGGGTGCT 55

RESULT 27  
US-09-252-991A-9308  
; Sequence 9308, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9308  
; LENGTH: 456  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9308

Query Match 3.4%; Score 17; DB 4; Length 456;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 TCGGCTGCGCGGGGC 363  
DB 138 TCGGCTGCGCGGGGC 154

RESULT 28  
US-09-060-756-680/c  
; Sequence 680, Application US/09060756  
; Patent No. 6183957  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/060,756  
; CURRENT FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 680  
; LENGTH: 507  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (various positions within the sequence)  
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"

US-09-060-756-680

Query Match 3.4%; Score 17; DB 4; Length 768;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SEQUENCE CHARACTERISTICS:  
LENGTH: 972 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US95-04801-2

Query Match 3.4%; Score 17; DB 5; Length 972;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 GCGCGGGGGCGCGTCC 370  
DB 954 GCGCGGGGGCGCGTCC 970

RESULT 33

US-09-252-991A-15003  
Sequence 15003, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 15003  
LENGTH: 999  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15003

Query Match 3.4%; Score 17; DB 4; Length 999;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCGGACGCGTGTG 237  
DB 281 CCGCGGACGCGTGTG 297

RESULT 34

US-09-252-991A-6751  
Sequence 6751, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6751  
LENGTH: 1062  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6751

Query Match 3.4%; Score 17; DB 4; Length 1062;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 ACCACTGGTGGGGCC 309  
DB 503 ACCACTGGTGGGGCC 519

RESULT 35

US-08-529-600D-1  
Sequence 1, Application US/08529600D  
Patent No. 5861285  
GENERAL INFORMATION:  
APPLICANT: Tadashi MATSUNAGA  
TITLE OF INVENTION: PROTEIN-BOUND MAGNETIC PARTICLES AND  
TITLE OF INVENTION: PROCESS OF PRODUCING THE SAME  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PILLSBURY MADISON & SUTRO, L.L.P.  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/529,600D  
FILING DATE: 18-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-248700  
FILING DATE: 16-SEP-1994  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1302 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1302  
US-08-529-600D-1

Query Match 3.4%; Score 17; DB 2; Length 1302;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 CCTTCCTGGCGCGGTG 279  
DB 971 CCTTCCTGGCGCGGTG 987

RESULT 36

US-08-973-275-5  
Sequence 5, Application US/08973275B  
Patent No. 5958706  
GENERAL INFORMATION:  
APPLICANT: MATSUNAGA, Tadashi  
APPLICANT: KAMIYA, Shinji  
APPLICANT: NAKABA, Kenryo  
TITLE OF INVENTION: FINE MAGNETIC PARTICLES CONTAINING USEFUL PROTEINS  
TITLE OF INVENTION: BOUND THERETO, PROCESS FOR PRODUCING THE SAME, AND USE  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: MATSUNAGA  
CURRENT APPLICATION NUMBER: US/08/973,275B  
CURRENT FILING DATE: 1998-02-09  
EARLIER APPLICATION NUMBER: PCT/Jp97/01043  
EARLIER FILING DATE: 1997-03-27  
EARLIER APPLICATION NUMBER: JP 8-97536  
EARLIER FILING DATE: 1996-03-28

Thu Jun 17 09:02:22 2004

EARLIER APPLICATION NUMBER: JP 8-146833  
 EARLIER FILING DATE: 1996-05-16  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: Patentin Ver. 2.0  
 SEQ ID NO 5  
 LENGTH: 1302  
 TYPE: DNA  
 ORGANISM: Magnetospirillum AMB-1  
 FEATURE:  
 NAME/KEY: gene  
 LOCATION: (1)..(1302)  
 OTHER INFORMATION: Genomic DNA sequence encoding mag A gene.

US-08-973-275-5

Query Match 3.4%; Score 17; DB 2; Length 1302;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 CCTTCTGCGCGCGGTG 279  
 Db 971 CCTTCTGCGCGCGGTG 987

RESULT 37  
 US-09-122-632-1  
 Sequence 1, Application US/09122632  
 Patent No. 6033878  
 GENERAL INFORMATION:  
 APPLICANT: Tadaishi MATSUNAGA  
 TITLE OF INVENTION: PROTEIN-BOUND MAGNETIC PARTICLES AND  
 TITLE OF INVENTION: PROCESS OF PRODUCING THE SAME  
 NUMBER OF SEQUENCES: 3  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: PILLSBURY MADISON & SUTRO, L.L.P.  
 STREET: 1100 NEW YORK AVENUE, N.W.  
 CITY: WASHINGTON  
 STATE: D.C.  
 COUNTRY: U.S.A.  
 ZIP: 20005-3918  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Microsoft Word  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/122,632  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/529,600  
 FILING DATE: 18-SEP-1995  
 APPLICATION NUMBER: JP 6-248700  
 FILING DATE: 16-SEP-1994  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1302 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1302  
 US-09-122-632-1

Query Match 3.4%; Score 17; DB 3; Length 1302;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 CCTTCTGCGCGCGGTG 279  
 Db 971 CCTTCTGCGCGCGGTG 987

RESULT 39  
 US-09-489-039A-4787/c  
 Sequence 4787, Application US/09489039A  
 Patent No. 6610836  
 GENERAL INFORMATION:  
 APPLICANT: Gary Breton et. al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 2709.2004001  
 CURRENT APPLICATION NUMBER: US/09/489,039A  
 CURRENT FILING DATE: 2000-01-27  
 PRIOR APPLICATION NUMBER: US 60/117,747  
 PRIOR FILING DATE: 1999-01-29  
 NUMBER OF SEQ ID NOS: 14342  
 SEQ ID NO 4787  
 LENGTH: 1587  
 TYPE: DNA  
 ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-4787

Query Match 3.4%; Score 17; DB 4; Length 1587;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 GCCGCCCGCGCGCGG 323  
 Db 273 GCCGCCCGCGCGCGG 257

RESULT 40  
 US-07-820-011A-1  
 Sequence 1, Application US/07820011A  
 Patent No. 5336615  
 GENERAL INFORMATION:  
 APPLICANT: Bell, Leonard  
 APPLICANT: Madri, Joseph A.  
 APPLICANT: Warren, Stephen L.  
 APPLICANT: Luthringer, Daniel J.  
 TITLE OF INVENTION: Genetically Engineered  
 TITLE OF INVENTION: Endothelial Cells Exhibiting Enhanced  
 TITLE OF INVENTION: Migration

Query Match 3.4%; Score 17; DB 4; Length 1587;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 GCCGCCCGCGCGCGG 323  
 Db 273 GCCGCCCGCGCGCGG 257

RESULT 40  
 US-07-820-011A-1  
 Sequence 1, Application US/07820011A  
 Patent No. 5336615  
 GENERAL INFORMATION:  
 APPLICANT: Bell, Leonard  
 APPLICANT: Madri, Joseph A.  
 APPLICANT: Warren, Stephen L.  
 APPLICANT: Luthringer, Daniel J.  
 TITLE OF INVENTION: Genetically Engineered  
 TITLE OF INVENTION: Endothelial Cells Exhibiting Enhanced  
 TITLE OF INVENTION: Migration

Query Match 3.4%; Score 17; DB 4; Length 1587;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 GCCGCCCGCGCGCGG 323  
 Db 273 GCCGCCCGCGCGCGG 257

RESULT 40  
 US-07-820-011A-1  
 Sequence 1, Application US/07820011A  
 Patent No. 5336615  
 GENERAL INFORMATION:  
 APPLICANT: Bell, Leonard  
 APPLICANT: Madri, Joseph A.  
 APPLICANT: Warren, Stephen L.  
 APPLICANT: Luthringer, Daniel J.  
 TITLE OF INVENTION: Genetically Engineered  
 TITLE OF INVENTION: Endothelial Cells Exhibiting Enhanced  
 TITLE OF INVENTION: Migration

Query Match 3.4%; Score 17; DB 4; Length 1587;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 GCCGCCCGCGCGCGG 323  
 Db 273 GCCGCCCGCGCGCGG 257

RESULT 38  
 US-09-252-991A-6874/c  
 Sequence 6874, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 6874  
 LENGTH: 1530  
 TYPE: DNA  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-6874

Query Match 3.4%; Score 17; DB 4; Length 1530;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 ACCACCTGGTGGCGGCC 309  
 Db 716 ACCACCTGGTGGCGGCC 700

RESULT 39  
 US-09-489-039A-4787/c  
 Sequence 4787, Application US/09489039A  
 Patent No. 6610836  
 GENERAL INFORMATION:  
 APPLICANT: Gary Breton et. al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 2709.2004001  
 CURRENT APPLICATION NUMBER: US/09/489,039A  
 CURRENT FILING DATE: 2000-01-27  
 PRIOR APPLICATION NUMBER: US 60/117,747  
 PRIOR FILING DATE: 1999-01-29  
 NUMBER OF SEQ ID NOS: 14342  
 SEQ ID NO 4787  
 LENGTH: 1587  
 TYPE: DNA  
 ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-4787

Query Match 3.4%; Score 17; DB 4; Length 1587;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 GCCGCCCGCGCGCGG 323  
 Db 273 GCCGCCCGCGCGCGG 257

RESULT 40  
 US-07-820-011A-1  
 Sequence 1, Application US/07820011A  
 Patent No. 5336615  
 GENERAL INFORMATION:  
 APPLICANT: Bell, Leonard  
 APPLICANT: Madri, Joseph A.  
 APPLICANT: Warren, Stephen L.  
 APPLICANT: Luthringer, Daniel J.  
 TITLE OF INVENTION: Genetically Engineered  
 TITLE OF INVENTION: Endothelial Cells Exhibiting Enhanced  
 TITLE OF INVENTION: Migration

Query Match 3.4%; Score 17; DB 4; Length 1587;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 GCCGCCCGCGCGCGG 323  
 Db 273 GCCGCCCGCGCGCGG 257

RESULT 40  
 US-07-820-011A-1  
 Sequence 1, Application US/07820011A  
 Patent No. 5336615  
 GENERAL INFORMATION:  
 APPLICANT: Bell, Leonard  
 APPLICANT: Madri, Joseph A.  
 APPLICANT: Warren, Stephen L.  
 APPLICANT: Luthringer, Daniel J.  
 TITLE OF INVENTION: Genetically Engineered  
 TITLE OF INVENTION: Endothelial Cells Exhibiting Enhanced  
 TITLE OF INVENTION: Migration

Query Match 3.4%; Score 17; DB 4; Length 1587;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 GCCGCCCGCGCGCGG 323  
 Db 273 GCCGCCCGCGCGCGG 257

RESULT 40  
 US-07-820-011A-1  
 Sequence 1, Application US/07820011A  
 Patent No. 5336615  
 GENERAL INFORMATION:  
 APPLICANT: Bell, Leonard  
 APPLICANT: Madri, Joseph A.  
 APPLICANT: Warren, Stephen L.  
 APPLICANT: Luthringer, Daniel J.  
 TITLE OF INVENTION: Genetically Engineered  
 TITLE OF INVENTION: Endothelial Cells Exhibiting Enhanced  
 TITLE OF INVENTION: Migration

Query Match 3.4%; Score 17; DB 4; Length 1587;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 GCCGCCCGCGCGCGG 323  
 Db 273 GCCGCCCGCGCGCGG 257

RESULT 40  
 US-07-820-011A-1  
 Sequence 1, Application US/07820011A  
 Patent No. 5336615  
 GENERAL INFORMATION:  
 APPLICANT: Bell, Leonard  
 APPLICANT: Madri, Joseph A.  
 APPLICANT: Warren, Stephen L.  
 APPLICANT: Luthringer, Daniel J.  
 TITLE OF INVENTION: Genetically Engineered  
 TITLE OF INVENTION: Endothelial Cells Exhibiting Enhanced  
 TITLE OF INVENTION: Migration

Query Match 3.4%; Score 17; DB 4; Length 1587;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 GCCGCCCGCGCGCGG 323  
 Db 273 GCCGCCCGCGCGCGG 257

RESULT 40  
 US-07-820-011A-1  
 Sequence 1, Application US/07820011A  
 Patent No. 5336615  
 GENERAL INFORMATION:  
 APPLICANT: Bell, Leonard  
 APPLICANT: Madri, Joseph A.  
 APPLICANT: Warren, Stephen L.  
 APPLICANT: Luthringer, Daniel J.  
 TITLE OF INVENTION: Genetically Engineered  
 TITLE OF INVENTION: Endothelial Cells Exhibiting Enhanced  
 TITLE OF INVENTION: Migration



Thu Jun 17 09:02:22 2004

TITLE OF INVENTION: and Plasminogen Activator Activity

NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Maurice M. Klee  
STREET: 1951 Burr Street  
CITY: Fairfield  
STATE: Connecticut  
COUNTRY: USA  
ZIP: 06430

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb storage  
COMPUTER: IBM PC XT  
OPERATING SYSTEM: PC-DOS/MS-DOS 2.10  
SOFTWARE: Displaywrite 3  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/820,011A  
FILING DATE: 19920106

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Klee, Maurice M.

REGISTRATION NUMBER: 30,399

REFERENCE/DOCKET NUMBER: 1B-101

TELEPHONE: (203) 255 1400

TELEFAX: (203) 255 1401

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1602 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: cdna to mRNA

HYPOTHETICAL: No

ANTI-SENSE: No

ORGANISM: Gallus, gallus

PUBLICATION INFORMATION:

AUTHORS: Takeya, Tatsuo

AUTHORS: Hanafusa, Hidesaburo

TITLE: Structure and Sequence of the

TITLE: Cellular Gene Homologous to the RSV src

TITLE: Gene and the Mechanism for Generating the

TITLE: Transforming Virus

JOURNAL: Cell

VOLUME: 32

PAGES: 881-890

DATE: March, 1983

PCT-US93-00445-1

Query Match 3.4%; Score 17; DB 1; Length 1602;

Best Local Similarity 100.0%; Pred. No. 70;

Mismatches 0; Indels 0; Gaps 0;

Matches 17; Conservative 0;

QY 437 GCACCGAGACCTGCGGG 453

DB 1149 GCACCGAGACCTGCGGG 1165

RESULT 41

PCT-US93-00445-1

Sequence 1, Application PC/TUS9300445

GENERAL INFORMATION:

APPLICANT: Bell, Leonard

APPLICANT: Madri, Joseph A.

APPLICANT: Warren, Stephen L.

APPLICANT: Luthringer, Daniel J.

TITLE OF INVENTION: Genetically Engineered

TITLE OF INVENTION: Endothelial Cells

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maurice M. Klee

STREET: 1951 Burr Street

CITY: Fairfield

STATE: Connecticut

COUNTRY: USA

ZIP: 06430

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb storage

COMPUTER: IBM PC XT

OPERATING SYSTEM: PC-DOS/MS-DOS 2.10

SOFTWARE: Displaywrite 3

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/820,011A

FILING DATE: 19920106

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Klee, Maurice M.

REGISTRATION NUMBER: 30,399

REFERENCE/DOCKET NUMBER: 1B-101

TELEPHONE: (203) 255 1400

TELEFAX: (203) 255 1401

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1602 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: cdna to mRNA

HYPOTHETICAL: No

ANTI-SENSE: No

ORGANISM: Gallus, gallus

PUBLICATION INFORMATION:

AUTHORS: Takeya, Tatsuo

AUTHORS: Hanafusa, Hidesaburo

TITLE: Structure and Sequence of the

TITLE: Cellular Gene Homologous to the RSV src

TITLE: Gene and the Mechanism for Generating the

TITLE: Transforming Virus

JOURNAL: Cell

VOLUME: 32

PAGES: 881-890

DATE: March, 1983

PCT-US93-00445-1

Query Match 3.4%; Score 17; DB 1; Length 1602;

Best Local Similarity 100.0%; Pred. No. 70;

Mismatches 0; Indels 0; Gaps 0;

Matches 17; Conservative 0;

QY 437 GCACCGAGACCTGCGGG 453

DB 1149 GCACCGAGACCTGCGGG 1165

RESULT 41

PCT-US93-00445-1

Sequence 1, Application PC/TUS9300445

GENERAL INFORMATION:

APPLICANT: Bell, Leonard

APPLICANT: Madri, Joseph A.

APPLICANT: Warren, Stephen L.

APPLICANT: Luthringer, Daniel J.

TITLE OF INVENTION: Genetically Engineered

TITLE OF INVENTION: Endothelial Cells

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maurice M. Klee

STREET: 1951 Burr Street

CITY: Fairfield

STATE: Connecticut

COUNTRY: USA

ZIP: 06430

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb storage

COMPUTER: IBM PC XT

OPERATING SYSTEM: PC-DOS/MS-DOS 2.10

SOFTWARE: Displaywrite 3

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/820,011A

FILING DATE: 19920106

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Klee, Maurice M.

REGISTRATION NUMBER: 30,399

REFERENCE/DOCKET NUMBER: 1B-101

TELEPHONE: (203) 255 1400

TELEFAX: (203) 255 1401

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1602 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: cdna to mRNA

HYPOTHETICAL: No

ANTI-SENSE: No

ORGANISM: Gallus, gallus

PUBLICATION INFORMATION:

AUTHORS: Takeya, Tatsuo

AUTHORS: Hanafusa, Hidesaburo

TITLE: Structure and Sequence of the

TITLE: Cellular Gene Homologous to the RSV src

TITLE: Gene and the Mechanism for Generating the

TITLE: Transforming Virus

JOURNAL: Cell

VOLUME: 32

PAGES: 881-890

DATE: March, 1983

PCT-US93-00445-1

Query Match 3.4%; Score 17; DB 1; Length 1602;

Best Local Similarity 100.0%; Pred. No. 70;

Mismatches 0; Indels 0; Gaps 0;

Matches 17; Conservative 0;

QY 437 GCACCGAGACCTGCGGG 453

DB 1149 GCACCGAGACCTGCGGG 1165

RESULT 41

PCT-US93-00445-1

Sequence 1, Application PC/TUS9300445

GENERAL INFORMATION:

APPLICANT: Bell, Leonard

APPLICANT: Madri, Joseph A.

APPLICANT: Warren, Stephen L.

APPLICANT: Luthringer, Daniel J.

TITLE OF INVENTION: Genetically Engineered

TITLE OF INVENTION: Endothelial Cells

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maurice M. Klee

STREET: 1951 Burr Street

CITY: Fairfield

STATE: Connecticut

COUNTRY: USA

ZIP: 06430

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb storage

COMPUTER: IBM PC XT

OPERATING SYSTEM: PC-DOS/MS-DOS 2.10

SOFTWARE: Displaywrite 3

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/820,011A

FILING DATE: 19920106

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Klee, Maurice M.

REGISTRATION NUMBER: 30,399

REFERENCE/DOCKET NUMBER: 1B-101

TELEPHONE: (203) 255 1400

TELEFAX: (203) 255 1401

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1602 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: cdna to mRNA

HYPOTHETICAL: No

ANTI-SENSE: No

ORGANISM: Gallus, gallus

PUBLICATION INFORMATION:

AUTHORS: Takeya, Tatsuo

AUTHORS: Hanafusa, Hidesaburo

TITLE: Structure and Sequence of the

TITLE: Cellular Gene Homologous to the RSV src

TITLE: Gene and the Mechanism for Generating the

TITLE: Transforming Virus

JOURNAL: Cell

VOLUME: 32

PAGES: 881-890

DATE: March, 1983

PCT-US93-00445-1

Query Match 3.4%; Score 17; DB 1; Length 1602;

Best Local Similarity 100.0%; Pred. No. 70;

Mismatches 0; Indels 0; Gaps 0;

Matches 17; Conservative 0;

QY 437 GCACCGAGACCTGCGGG 453

DB 1149 GCACCGAGACCTGCGGG 1165

RESULT 41

PCT-US93-00445-1

Sequence 1, Application PC/TUS9300445

GENERAL INFORMATION:

APPLICANT: Bell, Leonard

APPLICANT: Madri, Joseph A.

APPLICANT: Warren, Stephen L.

APPLICANT: Luthringer, Daniel J.

TITLE OF INVENTION: Genetically Engineered

TITLE OF INVENTION: Endothelial Cells

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maurice M. Klee

STREET: 1951 Burr Street

CITY: Fairfield

STATE: Connecticut

COUNTRY: USA

ZIP: 06430

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb storage

COMPUTER: IBM PC XT

OPERATING SYSTEM: PC-DOS/MS-DOS 2.10

SOFTWARE: Displaywrite 3

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/820,011A

FILING DATE: 19920106

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Klee, Maurice M.

REGISTRATION NUMBER: 30,399

REFERENCE/DOCKET NUMBER: 1B-101

ATTORNEY/AGENT INFORMATION:  
NAME: Wallen III, John W.  
REGISTRATION NUMBER: 35,403  
REFERENCE/DOCKET NUMBER: 19179  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-3905  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1700 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US95-04801-1

Query Match 3.4%; Score 17; DB 5; Length 1700;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 GGCGGGGGCGGGTCC 370  
|||  
Db 54 GGCGGGGGCGGGTCC 70

RESULT 45  
US-09-470-881-2  
Sequence 2, Application US/09470881  
Patent No. 6685938  
GENERAL INFORMATION:  
APPLICANT: CHERESH, David A.  
APPLICANT: ELICEIRI, Brian  
TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL FOR MODULATION OF  
TITLE OF INVENTION: ANGIOGENESIS AND VASCULAR PERMEABILITY USING SRC OR  
TITLE OF INVENTION: YES TYROSINE KINASES  
FILE REFERENCE: TSRI 651.2  
CURRENT APPLICATION NUMBER: US/09/470,881  
CURRENT FILING DATE: 1999-12-22  
PRIOR APPLICATION NUMBER: PCT/US99/11780  
PRIOR FILING DATE: 1999-05-28  
PRIOR APPLICATION NUMBER: 60/087,220  
PRIOR FILING DATE: 1998-05-29  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1759  
TYPE: DNA  
ORGANISM: Chicken  
FEATURE:  
NAME/KEY: gene  
LOCATION: (1)..(1759)  
OTHER INFORMATION: chicken c-SRC cDNA  
NAME/KEY: CDS  
LOCATION: (112)..(1710)  
US-09-470-881-2

Query Match 3.4%; Score 17; DB 4; Length 1759;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 GCACCGAGACCTGGGG 453  
|||  
Db 1260 GCACCGAGACCTGGGG 1276

Search completed: June 16, 2004, 19:32:35  
Job time : 70.3333 secs

TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9291

Query Match 3.4%; Score 17; DB 4; Length 1629;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 TCGGCCTGGCGGGGC 363  
|||  
Db 944 TCGGCCTGGCGGGGC 960

RESULT 43  
US-09-252-991A-9280/c  
Sequence 9280, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 9280  
LENGTH: 1641  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9280

Query Match 3.4%; Score 17; DB 4; Length 1641;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 TCGGCCTGGCGGGGC 363  
|||  
Db 852 TCGGCCTGGCGGGGC 836

RESULT 44  
PCT-US95-04801-1  
Sequence 1, Application PC/TUS9504801  
GENERAL INFORMATION:  
APPLICANT: Martin, Juan F.  
APPLICANT: Coque, Juan R.  
APPLICANT: Enguita, Francisco J.  
APPLICANT: Fuente, Juan L.  
APPLICANT: Lliarena, Francisco J.  
APPLICANT: Liras, Paloma  
TITLE OF INVENTION: DNA ENCODING CEPHAMYCIN BIOSYNTHESIS  
TITLE OF INVENTION: LATE GENES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John W. Wallen III  
STREET: P.O. Box 2000  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04801  
FILING DATE:  
CLASSIFICATION:

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:02:16 ; Search time 2074.33 Seconds  
(without alignments)  
10468.356 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325  
Perfect score: 501  
Sequence: 1 atgttctacacgtcggtgcac.....cccggtcgagggtcgtgc 501

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 15  
Total number of hits satisfying chosen parameters: 8137

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pt.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	501	100.0	53784	1	AMM223012	AJ223012 Amycolato
2	501	100.0	53789	6	A69720	A69720 Sequence 3
3	501	100.0	90445	1	AF040570	AF040570 Amycolato
4	28	5.6	69644	1	AY179507	AY179507 Streptomy
5	26	5.2	82746	1	AF453501	AF453501 Actinosyn
6	26	5.2	88421	6	AX417445	AX417445 Sequence
7	24	4.8	32870	1	AF007101	AF007101 Streptomy
8	24	4.8	347750	1	AP002998	AP002998 Mesorhizo
9	23	4.6	312050	1	MLCB1779	MLCB1779 Mycobacteri
10	23	4.6	312050	1	MLCB1779	MLCB1779 Mycobacteri
11	22	4.4	7887	6	BD262946	BD262946 DNA sequ
12	22	4.4	9858	1	AF321122	AF321122 Streptomy
13	22	4.4	9858	6	AX803758	AX803758 Sequence
14	22	4.4	18876	6	AF324838	AF324838 Streptomy
15	22	4.4	34869	1	BD262936	BD262936 DNA sequ
16	22	4.4	37856	6	AX024212	AX024212 Sequence
17	22	4.4	37856	1	AF322256	AF322256 Streptomy
18	22	4.4	39428	1	AX803750	AX803750 Sequence
19	22	4.4	61944	6	AX803750	AX803750 Sequence
20	21	4.2	144560	2	EX664737	EX664737 Homo sapi
21	21	4.2	189050	1	AL646066	AL646066 Ralstonia
22	21	4.2	219447	9	HUMFLNG6PD	L44140 Homo sapien
23	21	4.2	349305	1	EX640433	EX640433 Bordetell
24	20	4.0	5249	4	S67307	S67307 Ca(2+) - sens
25	20	4.0	5275	6	AR012622	AR012622 Sequence
26	20	4.0	5275	6	AR028465	AR028465 Sequence
27	20	4.0	5275	6	AR078217	AR078217 Sequence
28	20	4.0	5275	6	AR177781	AR177781 Sequence
29	20	4.0	5275	6	I75051	I75051 Sequence 1
30	20	4.0	11842	1	AE005053	AE005053 Halobacte
31	20	4.0	23451	1	AB070952	AB070952 Streptomy
32	20	4.0	28890	1	AF512431	AF512431 Saccharot
33	20	4.0	77292	8	EX842596	EX842596 Neurospor
34	20	4.0	102842	8	AF497474	AF497474 Aegilops
35	20	4.0	203050	1	AL646071	AL646071 Ralstonia
36	20	4.0	208524	1	AE016925	AE016925 Chromobac
37	20	4.0	299800	1	AP005040	AP005040 Streptomy
38	20	4.0	300900	1	AP005939	AP005939 Bradyrhiz
39	20	4.0	302300	1	AP005034	AP005034 Streptomy
40	19	3.8	558	6	AX756416	AX756416 Sequence
41	19	3.8	579	8	AF267643	AF267643 Zea mays
42	19	3.8	716	1	AF481794	AF481794 Streptomy
43	19	3.8	1150	8	AF079237	AF079237 Arundinel
44	19	3.8	1172	6	AR170948	AR170948 Sequence
45	19	3.8	1178	6	AR170983	AR170983 Sequence

ALIGNMENTS

RESULT 1

AMM223012

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AMM223012 53784 bp DNA linear BCT 09-FEB-1998  
Amycolatopsis mediterranei genes encoding rifamycin polyketide  
synthases, ORFs 1 to 5.

AJ223012 GI:2764760  
ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.

Amycolatopsis mediterranei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.

1 Schupp,T., Toupet,C., Engel,N. and Goff,S.  
Cloning and sequence analysis of the putative rifamycin polyketide

Pred. No. is the number of results predicted by chance to have a

synthase gene cluster from Amycolatopsis mediterranei  
 Unpublished  
 2 (bases 1 to 53784)  
 Schupp, T.  
 Direct Submission  
 Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma  
 AG, CH 4002 Basel, Postfach, Novartis K-681.344, 4002 Basel,  
 SWITZERLAND

## FEATURES

Location/Qualifiers

1..53784  
 /organism="Amycolatopsis mediterranei"  
 /mol\_type="genomic DNA"  
 /strain="LBG A3136"  
 /db\_xref="taxon:33910"  
 1336..15543  
 /function="CoA ligase, acylcarrier protein,  
 b-ketoacyl-ACP-synthase, acyltransferase, ketoreductase"  
 /note="ORF1  
 PKS modules 1-3"  
 /codon\_start=1  
 /transl\_table=11  
 /product="rifamycin polyketide synthase, type 1"  
 /protein\_id="CAAL1035.1"  
 /db\_xref="GI:2764761"  
 /db\_xref="GOA:O54666"  
 /db\_xref="SPTREMBL:O54666"

## CDS

/translation="MRTDLIKPLHVLHLENATRFAGKPAFADHRTVTYVDLEARTRR  
 LAGHLAGVRGDRVAICLNKRVSTVESYFALLRAGANGVPLNPGSATAELHEPLTD  
 SGATVYTDAAQARLRRLAPHVELLVGGDVPGEHSHYDELALSRPAEPAADDELDE  
 PMATVYTSQGTTRPGKGVSTORNCMSVASCVPFPLGSDQRVLWPLFLPHSLSHIA  
 CVLSATVAGSVIADGSSADDVMRLIEAESSTFLAGVPTTYHLVRAARQSGSLAP  
 LRIAGLAVGALGRSFEETFGVPLIDAYGSTTCGAI TNWPNRGAARQSGSLAP  
 PGVDRVDPDTGLDVPAGEGEVMSGNVLMGVHNSPEATAAARDQWERTGLDAR  
 RDDAGFTTGRIKELIKRGGANHPGEVAVLRTVDGVAAGVSGPHDTLGEVPA  
 VYIPGFTDPAALTEKREQSIANYKVPDRILLEVHAIPTASGKIRGRLTDEBAQLR  
 YAAETHEESVAAALRLARLSGLDERAQCELLEDLVRTQADVLIGQVPDGRRA  
 FDLGFTSLAI VLNRLNLTGLMPLPASAFDHTPTFAALAAVRAELGITQAVABP  
 VVADDPGEPIATVGMACRLPGGVASPEDLWRLVAERVDVAFSPFGDRGMDLSDIDPD  
 RERAGTSYVGGEFLHDAGEFDAGFFGISPREAVMDPOORLLLETSEWALENAGVDP  
 IALKGDTDGVFGIMGQGVGSNAVELSGFTVTGTGVASSVSGSVYVLEGPVATV  
 DTACSSILVAMHAAQLRQSCSMALAGVTVMATPGSFVSPFSRORALAPDGRCKAP  
 AAAADGTGSEGVVYRLSVARERHRLIAVLRGSVNDQASNGLTAPNGLSQQ  
 RVTRALAAAGALGVDDVVEAHGTGTGLDPTIEAQLLATYQERKQPLMLGSLKSN  
 IGHQAAGVAGVYKMWALRHETPLTLVDKPTLEVDWSAGAILLETLARAWPRNG  
 RPRAGVSFGVSGTNAHLILEAPAEPPVAPPELVPVLSARSTESLSGOAERLA  
 SLEGDVSLTEVAGALVSRALDERAVAVVAGSEEAATGIRALNTAGSGTSGKVYV  
 FPQGTOWAGMGRELLAESVFAERTAECAALAPWIDSLVDVLRGEGLDERVDLQ  
 PACFAVMGLAAVMSVGVPRDPAVGHSGQETIAACVSGALSLEDAAKVAALRSQAIA  
 AELSGRGMSVALGDDVVSRLVDGVEVAANGPSSVVIAGDAHALDATLEILSGEG  
 IRVRRVADYASHTRVEDIRDTLAETLAGISAQAPAVPYSTVTSEWVRDAGVDLGG  
 YWYRNLNNOVRFCGAATALLCOQHTVFVEVSAHPVTVPQLSELTGDAIGTLRREDGGL  
 RRLIANGSELFPVGGIDVDMTAMPVAGVDLPTVYAEHHPYLEPAEPASAGDPLLGT  
 VVSPGSDRLTA VQMSRRQAQWADGLVFNALVEANIKGLDLAGTPVVUGELVYDAP  
 VVLPFRGSRVQLVQMSRQRPRPIEVFSREDEPMTREHAHTGLAPAAAPEPPAA  
 GDATDVVAGLRDADRYGIHPALLDAARTVVGDDLLPSVMTVGSLLASATATVTP  
 TATGRLTDPAGQPLTVESVRGTPFVAEQGTIDALPRVDWPEIPLPTASTADFLPYE  
 ATSAEATLSALQWLADPATRLVAVTGDCTEPCAAAILWGLVRSASQSEHGTIVLADL  
 DTPARLPAVVASQEPQVRVGVASVPELRTVTPQOARPLDPEGVLTIGTGTGLGA  
 LTPARLPAHGVNHLVLSRRGEAPLEELTALGASVATAACDVADRAQLEAVLRAI  
 PASHPTVATHTAGVDGDDVTETLPDLATVRRPKVDAALDLELREADLAAFLVDF  
 SSAAGVLNPGCYAAGAAELDALAQRLNSLDLPVSIANGYATVMSGTEHGLDAD  
 LRRNRLGMSGLPADEGMALDIAATGTGTLVADKFDVAALRATAGAGGVPPLIRGL  
 APIPREAAATASITRELGLAETEQAAALLDVRRHAAEVLGHSGSAEVSHTGFKD  
 AGFDSLTAVELRNLAAATGTLTSPAMIFDYKPKPALADHLRAKLFCSAANRPAEIGT  
 AABEEPITAVMACRFGGVHSPEDLWRLVADGADATPEPPADRGMDTRLIYHDDPH  
 EAGTTPVTHAGVLDAAFGFISNEALAMDPOQRLLLETSEWLEFPAADLPTTL  
 AGODIGVAGVNGSHDYSNMRHRAAGVEGFRITGGASVLSGRVAYHFGVGEPAVDT  
 ACSSILVAMHAAVQALORQSCSMALAGVMMGTVEITFVFSRQGLAPDGRCKAFAD  
 GADGTGNSGEGVGLLVELLSPARQGHQVAVVSGVANSQASNGLTAPNPGSQOVR  
 IRKALAAAGVSDVDVAEAGTGTGLDPIEAALLATYQNRETPLNLGSKVNSLNG  
 HTQAAAGVAGVYKMWAMRHGVLPRTLHVDPRSSVVDMSAGVSELTLEARDWNSGHP  
 RRAAGVSFGIGTNAHVVEEVAAPITTPQEPAPFVLPVLSARTAGLRGQAGRLA  
 AFDGDRTDVRVPDAAVALATTQAQLDHRVAVVLSADRAQLCADLAAFGSVGVVTGTPVDG

## CDS

/function="b-ketoacyl-ACP-synthase, acyltransferase,  
 dehydratase, ketoreductase, acylcarrier protein"  
 /note="ORF2  
 PKS modules 4-6"  
 /codon\_start=1  
 /transl\_table=11  
 /product="rifamycin polyketide synthase, type 1"  
 /protein\_id="CAAL1036.1"  
 /db\_xref="GI:2764762"  
 /db\_xref="GOA:O52789"  
 /db\_xref="SPTREMBL:O52789"

/translation="MANQSRKNMSAPNIOVDALRSLKENVRVLRQENSALAAAAE  
 PVAIVSMCRYAGGIINGPDPFRVVSSEGADVTGPEPDRGWDVEGLYHPDPDPNGTYY  
 VREGAFLODAQDAGFGISPREALAMDPOQRQLLEVSWETLERAGIDPHSVRGSDI  
 GVYAGVYDQADPDLSGFEFMSLERALGTAGGVASRVAYITLGLBGPATVITMCGS  
 SILVATHLAQALRRGECSMALAGSTVMTATPGGVGFARQALAFDGRCKSYAAADG  
 SGWABGVILLERLSVARERHQVLAVIRGSVNDQASNGLTAPNPAQQRVIRKA  
 LASAGTSDVDVTVEGHGTGLVDPIEVQALATYQGRDPOQPLMLGSKSVGHT  
 QASGVAGVITKMYQSLRHQGLPATQVDVATPCQVDSAGAILLEASGRWPRNGHPR  
 GSVISFGASGTNAMLLEAPEDPVTPEAPATPGVPLVVSATATSLAAQALAEAV  
 GYSTLADVAGTIVSGAMLSERAVVAGSHEEATVGLRLARGESAPGLLSGRSGVP  
 KGVYVFPQGTQWAGKRELLDSSEVFAARTAECTALGRVWDVMDLTVLRGADLL  
 RDVYVQAPASFAVMGLAAVMSLGVPEPARLVGSGQETIAACVSGALSLEDAAKVA  
 LRSQAASIALSGRGMSVALGDDVVSRLVDGVEVAANGPSSVVIAGDAHALD  
 EALDLDQOVRIRRVADYASHTRVVAAEDALAEMLGGIQAQAPVPYVSTGTGW  
 VEDAGVLDGGYVTRNRRQVRFGFVAVBLIEQHRVFEVSAHPVLPQINELVDOTTE  
 AVYGTLRREDGGRLRLLSAELFVRGVTVDVWSGVLPSPSRVELFYIATFDHGYWLQ  
 PGTVYDVALRAGDELGLPGLLGAIVLPQSDGLVFTSRSLKSHPLWAGHAIAGVLI  
 MGSATDVALRAGDELGLPGLLGAIVLPQSDGLVFTSRSLKSHPLWAGHAIAGVLI  
 SMDGEDWTRHATGLLGSSTREPSRFPDPAAMPDAPEDIDVENFTDRLPDDDRSV  
 LFPWNGILVUHVAGSALRVVRVAPGGPDALTFOAADTGLVMTDMSLVRSVSAAL  
 CACQGRVWRGDEVPALVDDHREDAGKGLHALLDAALHTWAFANPDODRSV  
 LIPWNGILVUHVAGSALRVVRVAPGGPDALTFOAADTGLVMTDMSLVRSVSAAL  
 ETEAAGERDLSLFDVMDIEVPATETAHEAVLEAFGEAAPLELTSLVRSVSAAL  
 AADARLVVYTRGAVREMTDPAGAAVMGLVRAAQANPGRILIVDTDGDVPLGALAS  
 GEPQLAVRGNAFVPRLARATGEVPEAPVFSPEGTVLTGTGSLGLAKHILAVH  
 GVRPLVLSARGVAALVTLEQCATSVVACVSDRQDQVALLAEHRTPTGVHLA  
 GLDDGIVGAINRERLAGVPAPKVDVAQHIDELTRDLGLDGLAFVVFSSAALMSGAGGQ

NYAANAVALDGLMAGRAAGLPCVSLAWGLWEADGLTANLSATDOARMSRGVLPMT  
PAAELDFIDIGLAAEQALVPIKLDLRTLGQATAGGVEPHLLRGLTRGSRFRTTAA  
ASGGGLVKTAGRAPBEQAVLLGIVQARAALVGFNAPELAQGTGSGFSLGFTSLA  
AVELNRRLSAATGVLKLPATLVDFPTVALARHLREELGETVAGAPATPVTTVADAGE  
PIAVGMACRLPGVMSPDLMWMAEGDQMSFFDRGWDLDGLFDSDFPBPPTAY  
IRQGGFLHEALFDPGFFGLPREALAMPQORLLLEASWEALRAGIDPTKARGDAV  
GVFSVSHDYLESLSNMAPELEGFVTTATAGSVASGRVSYTFPGFPAVTVDTACSS  
SLVATHLAQALRQGCCTMALAGVAVMSPIGVICMSRQRMAGDPRKVAFAADGADG  
TVLSEGVTVLWELRVARERHVRVLAUVRGSAVNOQASNGLTAENGPSQORVIRSA  
LAGAGLQPSVDVVEAHGTGTALGEPTEAQLLATYTKGRETEPLWGLSKNSNHTQA  
AVSFGISGTAHLLLEAPDAADTAEDVDAVVPVVISARSTGLSAGRLAFLDGLD  
DVPLTRVAGNLLSTRATLADRAVVVAGSAREARAGLTALARGESAGLVTGTAMFGK  
TVWVFGQGTQWAGMRELLBASPVFRLEECALOPWIDWLSLDLWEGEGLDRV  
DVLQACFAVMVGLAAVMSVVPDPAVLGHSQETAAACVSGALSLEDAKVAVALRS  
QALAAELSGRGGMSQLSDEVAARLAPWAGRVETAAVNGPASPVIAGDAELTEAV  
EVLGRRVAVDYASHTRVEDIQTDLAETLAGIDAQAPVVPFYFSTVAGEWITDAGVVD  
GGVYRLNRQVGFPAVELIEQHGAFVEVSAHPVIVQVISELTDVAVTGTLRDD  
GGVRLTSMALFVRGVPVDMATMAPARVELPTVAFDQHFWSLSPVADAPALGL  
AGADHPLLGAVLPLDQSDGLVFTSRLSVRTHPWLADGVPAAALVELAVRAGDEAGCPV  
LADLTVKLLVLPESGRLRVQVIVSGRTVEVVSQLEGAEDWERNAGHLSATAPHE  
AFDFTWPPAGAQQVDGLWRGDEIFAVALPEELDAGAGIHPFLDAAVQVPLAD  
EQPAEWSLVADHAGSALRVLPVGGALQAADTGLVLTADSGVLAEGLSAGKTRAG  
SLVAVDWTSTIADSAVPAENIEVVEAFGEPELTLGRVLEAVQVTLVTAADDAVLVV  
TRGAVRTPDAGAAVGLVRAAOAENPGRIFLIDTDEIIPALTDPEPEIAVRGKFF  
VPRITRAPSGAAVFRDGTVLISGAGALGSLVARLVRHGVKVLASRRGRDADG  
VADLVADLAADVAVVAVDSDRAQVALLDEHRTVAVHTAGVIDAGVITLDRILA  
TVFAPKVDVADVHLDTRDLDLDAFVSVSVFVMSVAFMGSAAYAAAFLLGLMANRR  
AAGPLGLSLAWGLDQSTGMAAGTDEATARMSSRGGIQTQEGMDLFDALUSSAE  
SLVPAKLDLGRVADAAAGVPHMLRGLVRAGRAQAAASTVDNGLAGLALAPA  
DQTLILLDLVRAQAVNLGHADAASVAVDTAFKADGFDLSLTAVELNRMRATATGLKP  
ATLVDPVNPOLLARHLDELGGAAQPTVTAABKADLDEPIALVGMACRLPGGVAGP  
EDLMRLVABAGDVSSPTDRGWDTSILYDPAAPPKTYTRHGGFLHEAGLFDAGFV  
GISPREAVAMPQORLLLEASWEAMEDAGVDPLSLKNDVGVFTGMFGQGVVAPQDSV  
VTPLEGFAGTGGSSVASGRVSVFVFGEPATVIDSSSSSLVAMHILAAQSLRQGC  
SMALAGATVMANPGAFVFSKORGLAVDRKCAFAAADGTGWAEGVGVVILERLSV  
AREGHTIRLIVLRASVNOGNSGLTAPNGPSQORVIRRALVSAGLAPSDVDVVEAH  
VLPPTLHVDRPTPEVDSAGAVELLTEAREWPRNRPRRAGVSARFVSGVTWHLLEE  
APAEFPVPTPEVPLVPVVSARSASLAGOAGRLAGVADSLASLAGARALVTRNRAAL  
TERAVMVVGSREEAVENTLEALARGEDPAAVVTGRAGSPKGLVWVFPFGQSQWIGMRE

Query Match 100.0%; Score 501; DB 1; Length 53784;  
Best Local Similarity 100.0%; Pred. No. 1.3e-233; Indels 0; Gaps 0;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTCTACAGCTCGGCGACGACCGGGCGGCCCAAGGGCGTGTGTCACGAGCGCAAC 60  
DB 1825 ATGTTCTACAGCTCGGCGACGACCGGGCGGCCCAAGGGCGTGTGTCACGAGCGCAAC 1884

QY 61 TGCCTCTGCTCGCTTCTGCTACGTCGCTTCCCGGGTTCGCGACGAGCGCG 120  
DB 1885 TGCCTCTGCTCGCTTCTGCTACGTCGCTTCCCGGGTTCGCGACGAGCGCG 1944

QY 121 GTGCTCTGCGCGCTCGCTTCCACAGCTTTCGACATCGCTTTCGCTTTCGCGC 180  
DB 1945 GTGCTCTGCGCGCTCGCTTCCACAGCTTTCGACATCGCTTTCGCGC 2004

QY 181 ACCGTGTCGGGGCGAGAGCTCGAGCTTCGCGAGCGGCGAGCTTCGCGACGAGCGG 240  
DB 2005 ACCGTGTCGGGGCGAGAGCTTCGCGAGCTTCGCGAGCGGCGAGCTTCGCG 2064

QY 241 CTGATCGAGCGGAGAGCTCGAGCTTCGCGAGCTTCGCGAGCTTCGCGAGCTTCGCG 300  
DB 2065 CTGATCGAGCGGAGAGCTCGAGCTTCGCGAGCTTCGCGAGCTTCGCGAGCTTCGCG 2124

QY 301 GTGCGGGCGCGCGCGAGCGGCTTCGCGAGCTTCGCGAGCTTCGCGAGCTTCGCGGG 360  
DB 2125 GTGCGGGCGCGCGCGAGCGGCTTCGCGAGCTTCGCGAGCTTCGCGAGCTTCGCGGG 2184

QY 421 ATCGAGCGCTTACGAGCGAGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAGAG 480  
DB 2245 ATCGAGCGCTTACGAGCGAGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAGAG 2304

QY 481 GCCCGCGCTCGAGGGCTCGTGC 501  
DB 2305 GCCCGCGCTCGAGGGCTCGTGC 2325

QY 421 ATCGAGCGCTTACGAGCGAGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAGAG 480  
DB 2245 ATCGAGCGCTTACGAGCGAGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAGAG 2304

QY 481 GCCCGCGCTCGAGGGCTCGTGC 501  
DB 2305 GCCCGCGCTCGAGGGCTCGTGC 2325

RESULT 2  
LOCUS A69720 53789 bp DNA linear PAT 07-MAY-1999  
DEFINITION Sequence 3 from Patent WO9807868.  
ACCESSION A69720  
VERSION A69720.1 GI:4774328  
KEYWORDS unidentified  
SOURCE unclassified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 53789)  
AUTHORS Schupp, T., Toupet, C. and Engel, N.  
TITLE RIFAMYCIN BIOSYNTHESIS GENE CLUSTER  
JOURNAL Patent: WO 9807868-A 3 26-FEB-1998;  
Ciba Geigy AG (CH)  
FEATURES  
source 1. 53789  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"

ORIGIN  
Query Match 100.0%; Score 501; DB 6; Length 53789;  
Best Local Similarity 100.0%; Pred. No. 1.3e-233;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTCTACAGCTCGGCGACGACCGGGCGGCCCAAGGGCGTGTGTCACGAGCGCAAC 60  
DB 1825 ATGTTCTACAGCTCGGCGACGACCGGGCGGCCCAAGGGCGTGTGTCACGAGCGCAAC 1884

QY 61 TGCCTCTGCTCGCTTCTGCTACGTCGCTTCCCGGGTTCGCGACGAGCGCG 120  
DB 1885 TGCCTCTGCTCGCTTCTGCTACGTCGCTTCCCGGGTTCGCGACGAGCGCG 1944

QY 121 GTGCTCTGCGCGCTCGCTTCCACAGCTTTCGACATCGCTTTCGCTTTCGCGC 180  
DB 1945 GTGCTCTGCGCGCTCGCTTCCACAGCTTTCGACATCGCTTTCGCGC 2004

QY 181 ACCGTGTCGGGGCGAGAGCTTCGCGAGCTTCGCGAGCGGCGAGCTTCGCGACGAGCGG 240  
DB 2005 ACCGTGTCGGGGCGAGAGCTTCGCGAGCTTCGCGAGCGGCGAGCTTCGCG 2064

QY 241 CTGATCGAGCGGAGAGCTCGAGCTTCGCGAGCTTCGCGAGCTTCGCGAGCTTCGCG 300  
DB 2065 CTGATCGAGCGGAGAGCTTCGCGAGCTTCGCGAGCTTCGCGAGCTTCGCGAGCTTCGCG 2124

QY 301 GTGCGGGCGCGCGCGAGCGGCTTCGCGAGCTTCGCGAGCTTCGCGAGCTTCGCGGG 360  
DB 2125 GTGCGGGCGCGCGCGAGCGGCTTCGCGAGCTTCGCGAGCTTCGCGAGCTTCGCGGG 2184

QY 421 ATCGAGCGCTTACGAGCGAGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAGAG 480  
DB 2245 ATCGAGCGCTTACGAGCGAGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAGAG 2304

QY 481 GCCCGCGCTCGAGGGCTCGTGC 501  
DB 2305 GCCCGCGCTCGAGGGCTCGTGC 2325

RESULT 3  
LOCUS AF040570 90445 bp DNA linear BCT 16-APR-2001  
DEFINITION Amycolatopsis mediterranei rifamycin biosynthetic gene cluster.  
ACCESSION AF040570 AF040571  
VERSION AF040570.2 GI:12331604  
KEYWORDS  
SOURCE  
ORGANISM  
Amycolatopsis mediterranei  
Amycolatopsis mediterranei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.  
1 (bases 1 to 90445)  
August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R., Yu,T.W.,  
Taylor,M., Hoffman,D., Kim,C.G., Zhang,X., Hutchinson,C.R. and  
Floss,H.G.  
Biosynthesis of the ansamycin antibiotic rifamycin: deductions from  
the molecular analysis of the rif biosynthetic gene cluster of  
Amycolatopsis mediterranei S699  
Chem. Biol. 5 (2), 69-79 (1998)  
98174059  
9512878  
2 (bases 1 to 90445)  
Kim,C.G., Yu,T.W., Fryhle,C.B., Handa,S. and Floss,H.G.  
3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the  
formation of the precursor of m<sup>7</sup>N units in rifamycin and related  
antibiotics  
J. Biol. Chem. 273 (11), 6030-6040 (1998)  
98165773  
9497318  
3 (bases 1 to 90445)  
Yu,T.W., Muller,R., Muller,M., Zhang,X., Draeger,G., Kim,C.G.,  
Leisner,E. and Floss,H.G.  
Mutational analysis and reconstituted expression of the  
biosynthetic genes involved in the formation of  
3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin  
biosynthesis in amycolatopsis mediterranei S699  
J. Biol. Chem. 276 (16), 12546-12555 (2001)  
21201076  
11278540  
4 (bases 1 to 90445)  
August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R.,  
Hutchinson,C.R. and Floss,H.G.  
Direct Submission  
Submitted (30-DEC-1997) Dept. of Chemistry, University of  
Washington, Box 351700, Seattle, WA 98195-1700, USA  
5 (bases 1 to 90445)  
August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R.,  
Hutchinson,C.R. and Floss,H.G.  
Direct Submission  
Submitted (19-JAN-2001) Dept. of Chemistry, University of  
Washington, Box 351700, Seattle, WA 98195-1700, USA  
Sequence update by submitter  
On or before Jan 22, 2001 this sequence version replaced  
gi:2792340, gi:2832263.  
FEATURES  
source  
1..90445  
/organism="Amycolatopsis mediterranei"  
/mol\_type="genomic DNA"  
/strain="S699"  
/db\_xref="taxon:33910"  
423..1451  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAC01706.1"  
/db\_xref="GI:2792310"  
/translation="MAGTGGVGSRRVVKTLRQLEVALIGAGLIARLHLEAWLGGAAV  
RVYSDGRSRELAEFGAAGACSLBLEDGADAVDICTPTASHEIALTAIAAGVGW  
CEPLAASAEAEELVTAARAGRLVIAHDVPAAPYAEVLHSLVAGRLGEGALGRF  
SFAYHPRTWTHASARGSGILLTQLLHGADLAHWFVGDVVRVHACVQGDIAPEEG  
AVATGTAVLTHASCAISQVSRWTATPRPVRVAFVHSGTGSVYDSEWQEVVD  
GGAGNFAYGGSPVDFTEMTREFAFAAGGPEPRICAKDALAAVRIIHAHAESAWTGRAV  
ELPVRGNA"

CDS  
1448..2215  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAC01707.1"  
/db\_xref="GI:2792311"  
/translation="MKVAILSTPQAVAGALRGLPDVEVVAASWDAPEPVRQAAEAG  
ARVICEYPPAAKFTDLKAMDADAGRLTEFASPGCHGAFVVRKGIADGIGELTIVL  
GSVATSDVGLGAAAPYLLDLADAVLGGEPAAQVYACQVNIIVLSGIGESAULTVYR  
SQVASDFCRHGSATGLPAVTFIGDSQSVQIDAGPQLGGGERPELGGEDLEALMLKD  
FLGADGPGPGQALRTFRIIQAYESAHTGQFPVDL"  
2353..2619  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAC01708.1"  
/db\_xref="GI:2792312"  
/translation="MQDSVLGDPKQVEAMSVTVQKIALSGGPAELAPAHRSVDSG  
ALENLKIRHAGYEHFVHGGEFQAVDGCIDIAVFRWSHRTKIAE"  
2652..3842  
/note="ORF0; similar to Streptomyces Sp. Chop"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450 monooxygenase"  
/protein\_id="AAC01709.1"  
/db\_xref="GI:2792313"  
/translation="MTDAISFEVPMDRTDKEDPPAVFDSLRREERPLAKMVPDGHVGM  
LVSSEYELVRELSDLRFSHSCVEGHPFVTHQGVITHEPLIPGMFTHMDPPEHTRVK  
ILTGFTVRRASRLIPRAEAAEQIEVMRAGAPADVMDFAKPLVRLMLGSLGVL  
YEERDRVPVATLLHDAADPAEAAAYEVAGKFDEVIERRRQPODDLLISLVED  
LTQELRNIVTLLEAGYETTCAGALATVFALLHTDQLAAAEPEKDLDALELLR  
YLTVNQHYVTRTALEDVKGELIKKGDVTVLSLPAARNPDAKFGPAEIDIDRTSG  
HVAFGFIHQCLGNLARIELRAGFTALLRAFPELLRAVPADEVPLRLKLSGVFSVKKL  
PVSA"

gene  
3883..18090  
/gene="rifa"  
3883..18090  
/gene="rifa"  
/note="Rifa"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase"  
/protein\_id="AAC01710.1"  
/db\_xref="GI:2792314"  
/translation="MRTDLIKPLHVALLENATRAGKPAFADDDHRTVTYDLEARTTR  
LAGHLAGLVGRHGRVAICLGNRVSTVESYFAILRAGVGVPLNGSATALEPLTD  
SGATVVTDAQAARLARLAPHELVITGDDVPEGAHSYDELAGSEPAPEADDDLELDE  
PAMFYTSGTGRPKGVVSTQNCILMSVASCYFPFGLSDQDRVLPPLPFLSHSLIA  
CVLSATVGVASRIADSSADDVMRLIEASTFLAGVPTTYHLHRAARGFSAPS  
LRLGAGAVLGAGLSEFEETGVLIDAYGSTETCGAITMNPDPGARVSGCSGLAV  
FGVDVVDPDTGLDPAGEGEVWVSGNVLGVHNSPEATAAMRGVPHDTLGEVPA  
RDDAGYFTICRIKELIIRGGANLHPGEVEAVLRYDGVADAAGVGVPHDTLGEVPA  
YVIPGTGFDPAALIEKREQLSAYKVPDRILEVAHPTASGKIRRGILLTDEPAQLR  
YAATHEQSHADENSVAALRLGLDERAQCELLEDLVRTAADVLIGQVPGRA  
PRDLGTSIAIVELNRNLTEHTGLWLPASAVDFHTPAALAAARVSGWDLSDLPD  
VVAADPGPIALVGMACELPGVASPELWRLVAERVDVASEFFPDGRGMDLSDLPD  
RERAGTSIVGGFLHDAGEPAPELEGFTTGTGVASVAGSVSYVYLGLGPAVTV  
TALKGTDGTFVGLMGQGVGAVAPALEGFTTGTGVASVAGSVSYVYLGLGPAVTV  
DTACSSILVAMELAAQALQGCESMALAGGVVMTAPGSFVSFQRALAPGRCKAF  
AAAADGTGWSGVGVVLERISVAREGRHILAVLRGSVNDGASNLATPANGUSQQ  
RVIREALAAAGLAPSDVDVVEAHGTGTTGLDIEAQAALITGYQERKQPLWLSKSN  
IQRVIRVANDYASHTRVEDIRDTLAETLAGISQAQPAVPFVYTSVSEWRDLRGLDGG  
IELSGEGMAVALGDDDDVSRVLDGVGEVAAVNGPSSVVIAGDAHALDLELSDGG  
PACFAMVGLAAVWESVGRVDPVAVGSHSQGEIAAACVSGALDEDAKVALRQSAIA  
FPGQGTQWAGMGRLELAEFPAERIAECAALAPWIDMSLVLDVLRGEGDLGRVDVQ  
SILGEDVSLTEVAGALVSRVAVDIDBRVAVVAGSREAVTGLRALNTAGSGTQKVVW  
RPRRAGVSFGVSNHLLILEAPAEPPVAPPELVPLVVSARSTESLSQAERLIA  
IGHAQAAGVAGVIMVQALRHETLPTLHVDPKPTLEVDWSAGATELLTEARAWPRNG  
RRLIAMSGLRFGDIDVMTAMVPAAGWDLPTIAREHRLWLEPAPASAGDPLLGT  
YVNRNLNRVFRGAATALIEQGHTVFVVSAPHTVQPLSELTGDAICTLRREDGGL  
RRLIAMSGLRFGDIDVMTAMVPAAGWDLPTIAREHRLWLEPAPASAGDPLLGT  
VYIPRGSGLRLTAVQWRRRAQWADGLVFNALVEAAILGLDLAGTVPVGLVLDVAP  
VVLPRGSEVOLIVCEPGEQRRRPIEVFSREADEPWTBHAGTLPAPAAAEVPERAA  
GDAITDVTAGLRDADRYGHPALLDAAVTVTGDDLLPSVWTVGSLASGATVTVTTP







```

/product="putative secreted protein"
/db_xref="AA006910.1"
/db_xref="GI:28192595"
/translation="MDKRMGHRRIQTPRTTLATRAVLAAGVLTPTIASAGSAHAA
TPQAICISDRDELAKUSEDINSALEGSAAITSLHDRTNTTCTLDADHFDAS
TVKVTVLISLDAQKONKALTOEEDKHATAMITSDNATLAKQLGADKINGFLQ
AAGMTITLDSCHWGLQITANDDEKLLQLVHTNPVLSDDSRAYILKLTAEVIPSQ
RWGTAPAGSDAQVHKNGLERATNGWVHSLGATGCGDHDYTTITVLSQDNATWDDG
IANEGIAVHENINAPVSSAQ"
complement(1945..2736)
/notes="similar to S. coelicolor SCF1.09 (SC0267); ORF18"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="AA006911.1"
/db_xref="GI:28192596"
/translation="MTIRPVAGSLDVRVGMGTPLVFAHYMGSAADTWNGVIGHL
PQCATVFDQRMGSRALPGYHLIDQLADDLVRIEVRVSPFVLVGHSMGKAGQ
LVAAAPAGLVGLVAPAPQPPAMPTVEYRQGLSHAYDSVETGHALDHVITATPL
PEAVRATARDSLAAGAEARQEWPLRGVIARDITDAARRIEVPTVLAENDVVEPPI
LRSHLPHIPLATLITVPDAGHLLPAEAPGAVARALGDFLTATQS"
2830..3774
/notes="Arac family; similar to S. coelicolor SC5A7.14
(SC6644) and SCF1.08c (SC0266); ORF19"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AA006912.1"
/db_xref="GI:28192597"
/translation="MENGSAVRQLRVLPAVGSAHYGEVLDFAALRSMQDTRRTQPO
RPFHFVALPGSGTGSHEADPHNYRLGEGGAVIRPQWHRWSDIDACDGLILFRP
FLSGTASEATAPACWHLQRLSLALIAHLGRSESTAVHTPLASPVLLSHLLAA
LILRALGTPSPGSPGSRPTEVFRAVRAAEERTDWHVADYARALGYDVRTIT
RTAAAGTGAFLDQRLLEAKRLAHTDLPVSGCARLGRDVCNFTFFRRQAG
LPPAAARAAYSTAGAQG"
complement(3824..4585)
/notes="TetR family; similar to S. coelicolor 2SCD46.08
(SC4194); ORF20"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AA006913.1"
/db_xref="GI:28192598"
/translation="WLSVNTVTPSHLNAPTRLTRGKYPGAEIYQGFSAFVADVGK
PGLRERKQTHAAVDAASIALEFYGNOVSVAQVAAEAARVXRTLPAYEPTKEDLV
VRLADHETASVRAPPTAPLALREHLSELRDEPITGLNDQPEVRLRHM
LRLASVARMAGFKTGABRALTEALRETPELTARLAAQIVAVHWAQAQDNARL
AYGFADERYAGVADMEHAFALLENGRLHLPFRQ"
4569..6065
/genes="gdmI"
/notes="putative FAD-dependent monooxygenase"
/codon_start=1
/transl_table=11
/product="GdmI"
/protein_id="AA006914.1"
/db_xref="GI:28192599"
/translation="MLTSTTVVAGAGATGLMAYELALAGVETLVLEKLPIQOQ
VKGQIQRTEALLSRGELLEPLMRLAIRAPDVPFGSFGALPVLDCAPWTEHPPIG
IPQWIEVEERATAGARVLRGTAVSGVADDDGVVTVADGLRARHLYLVACDGGH
STVKILGLPFRGATHPAVLADIRLSAVSLVPRQMSLMTTRHARGYWSMLVPL
GDRYRFTGHQAQDARTDPTVTHEEIAAALQAVYPTTLGADVNSRFSATROL
EYHRTGRVLPAGDAHIHPPLCAQGLNIGVQDALNWKLAALVLODRAPNGLLSVHA
EHPVAAQVHTSQAQVLAISNPSEDVAALRDIPTDLRLPDRNRHLAGLMSGLSLR
YDLPDGHPTGRIPDADLVETGTITRLSLFGSSHANLLDLAGAVPADLPPLPRVDL
VRATCADDMAALIRPDGYVCWATDTSAAACDITLLAALTGLAKRP"
complement(6107..6577)
/genes="gdmX"
/notes="similar to JdxX and MmyY"
/codon_start=1

```

```

/transl_table=11
/product="GdmX"
/protein_id="AA006915.1"
/db_xref="GI:28192600"
/translation="MTATADYDSLARRLRALEDKALRALMLRGHRLDRKQWOTWIE
CWAEADVLEFGWEKIHGKEAVRAKVEEVSQMHILNMHVVDGEDRATGIMY
WFAVATPGSTSSPYSMGPGYDWEFRRPGDWLLVRQLGVWMTDGEOTLRAFA"
7208..27736
/genes="gdmA1"
/notes="polyketide synthase; modules 0-3"
/codon_start=1
/transl_table=11
/product="GdmA1"
/protein_id="AA006916.1"
/db_xref="GI:28192601"
/translation="MLRTDILIRPPELLRANADRFQDKPACSDGHRITVSHAELERRTR
RLAGHLGLRLHPGDRAMICLGNRVEMVESYFGLRVANGVAVPNRSTDAEUSYLLA
DSGARLVLDVAHADQFGLRQFPELRVVVSGDGLPKGFATFEPDTEPQLAARD
DGLDILTWLITSGTGLPKGVLSQTNCLWSLAACYVPVTLGAEDRVLWPLPLFH
SUSHVCLLAATAVAGASTRIYDGVSTADVLDALEERSSTFIAGVPTLYHHLIABAER
DFAATPELRTALVGGAVATADLVRSFESAFGVPLVDAYSGTSTCGATANWPTGPRVEG
SCGLPVEGLTVRLVDEDTVDVPAGEGFEFWGPNMAGIYHNGEATASALRDXHYR
TGBLGRDRAGCTVTGRKELVIRAGENIHGEVEAVLRIETARVPTASGKITRPHVLE
GEVPAFVVPDGPDPSSALATCERRLSYFKVPEEIEIETARVPTASGKITRPHVLE
LPLARAGDQYDLSLLRDWVPSALPDAPACTGTWALADADALGLAVGLRAAGVDA
RVGVEPVEPVADSVAGLVAGSVADLAGDDGAAPDVVVVTPVAGLVDETGADGAV
TVGERADRLAARLGAWLADRLLAGTTFVVAATGAVATGAEDAPEPLSALWGVRSLL
QAAYPQRLTIVDVLGDAGDAGDAGDEGREALLRAGVGGHQAIRGGLVPLPKL
TRISFPAEPQPPATLDAGLVITGTGTGTALARHLVAYGARNLLLSANGLPPE
AANTLRELRADGAOVSMAVCDPADRTALDSVLDAQAPVTAHVHIEEPERSLDTFS
LRAMTHRETRGAAAFVYVTSAGVLSGPGEPDRAADQCEALVRRRALGLGG
LALAWGLGEGHTGAPVAGAVLPEALALFDALTAGCQGLVILLRPSGTGPGCEPV
AVLRHVDAPSPADSEPAVAFRRNLAATCARLAATVVDHPTTGAARHLVRLEITGR
PVAGQAQFASGFTSLTAVARNLNLPEVATPEDLWRLVADGRDAIAGTFEDRGWL
RGVQAPVARGVSDPEVAVIYVAMGCHLPGEVATPEDLWRLVADGRDAIAGTFEDRGWL
AGLFDSPDAGVKSIVREGGFLTDAGGFDAAFFGISPREALAMPQQRLLLETAWET
ENAGIDPGLSGHTDVGVSVMVHDYGDAGTAAEGLEHGLVGSAGSVSVGRVAYAL
GLTPAVTVDTACSSSILVLAHVQAVRTGECSLAGAGVAVMSRPTSFIFRSQRL
APDRCKSFAGADGTNWSGVGLVLLERLSDARRNGHEVLAVVRGTAVNQDGSNGL
TAPNGSQERVIRQALANAGTVADVDVAEHGTGTSLGDPIDEAQALLATYQGRPEP
OPLWGLSKNIGHAQAAGAAAGVYKVMQMRHGVLPKTLHADPTSKDWTSGAVSL
LSEARPEPTCHPERAGISFSGVSTNAHVLEQALEAAAPATQASDAGAPGLVATG
GVVFWLSAKTIPALRAQARLUSHLESGSDANVDVWLSATTALEHRAVILATD
AEGMWALAEARGEPOLNVITGOTDGTGKTVPIPGCGCAQWGMGQALLATQSVFAA
RLRCCALAPYTDWLSLIDVITGTPDAPSLERVDVVPQPAFVAVVSLAALLNQSGVHP
DAVIGHSQGEITAAACVAGHLTNAKIVTLRSQIIAHLGAGHGMMLATPDIIDL
TNWEGKLIWAARNSPNATVAGDTDALHQLHTHYTDQGTARAIIPVDYASHTGVDTI
KNQQLQDVLGDTLEBGTIPMLSTVTGOWIEPNTVGDSSVYVYRNLRQTVQFERTIHLAD
QGYRTYIEISPHVLTALQETLEANDPTNTIVTGLRRDDDTPTRLTLNLHLITN
GTPVNWPTLFTGTQETRIPLPTYPQHHHYVLPNTSAGDVSAGVLCQGTGHPLAGAVV
SVPTDGVLLTGQLSVATHPLADHAVSGTVLLPGTAMAEALIRAGDEDTPTLEELV
IGQPMWLPEDGALHVQVLVGVDEGRGRIYSPRDAQOEQEWLHSHAGSTLATOQDGS
AEGMGEMNGMPWPPGVEPIALDDFYDDLAQAGVEYGFAPRGLKAVWRKDDGEVFAEA
LPEEDTVAGRFGIHPALLDAALHASFVCPVPPAGQOTLLPFVNGVRLLAAGATVRV
BARATGDSFTTSLSDTSGSVASVDSVLRALISPEOLAAASGAGSADALFTLDWT
EHPATGTEVSWATLGDATDVAHDVADIALAGEPDPGAVADTAAWAGDTCGLPARAR
DLAARLALVQRVGRPELADVLLTRGVASHVDHTDPAADAAINGLVRSAQSE
HFGRTALVDVTDVDSREALPEAVAAGEPQVALRRGLLWVRLVRSQGLAUPAHEHWYL
DVSEKSNGLVLRDPDAPATPLATQVRIEYRAAQONFRDVLVLRGWSVATGAGVGA
GAGVYEVSGVGEGLAVGDRVNGLFRSGFIADARTVADIPPEGWSVATGAGVFA
YLTALIGLDRGTQVPGGTVPFVAAGVGMAAQLARHFGATVATVAHPSKHVLTA
LGVPEGHLLASRLDGLFASAPALVDVNLSTGEYVDSILGLLGTGRFVEMKNDIRD
PASVAAAHGVGYQAFDGGDAPDRIRRELLABELVELPEAGRIEPLFIHWNVTOAT
HFGRTALVDVTDVDSREALPEAVAAGEPQVALRRGLLWVRLVRSQGLAUPAHEHWYL
ROGPDAPAGTDLTLELISGATVITACDTRDQDLAALADIPAAHPLTAVIHTAGA
LDDGVLTALTPDRLDVTRPKVDVATHLHLTRDQDLAALFVYSSAAGTLAGAGQAN
AAANAFILDAFAQRHARHARPATSLAWGIMSDTSLTSTMDATDVRTRRAGVGMNDA
LSGQLAGLSPAGQREFLNLVRAEAAVLAHAGPEALEPTVAFKEMGFDSLSLTAELRN
EALRVFDTLGRSRPALVAAKIDLTALRAPDAELSPILRGLARPARRTAARPAAGG
RINAATGLRLPATLLDHPTELLFELTGGGFPAPAAAAPVTVRAAADPFIAPVVA
MSCRLPGGVTDPTDGLWNLLLEERDGIADPRDRGWDLEALFDADPRDSGTSYVLRGGF

```



/product="RNA polymerase sigma factor"  
/protein\_id="AM54109.1"  
/db\_xref="GI:21449377"  
/translation="MWIGYSDGRRRRPVRWGVTEATFTWEDDIKLAATLRDAEA  
ANADSTRALKAIGRVLITADQVEIARRVEAGLYAEHLIDLPPLPAARGSRPVELR  
RVVRQQRKHLNANLRLVVSIAKRTYKGMALLDVIQENGLVRAVEKRPYAKG  
FKFYSYTWIRIQATRALADQSRTIRLPHVVEVVTKLARIERDLNREAEAE  
VAVELDVPEKVELROHARTVSLDQITGAEDTSIGDLIADENAVLEAVSALL  
SADLRVLETLTPREAGVMRYGFPDGVHLIDDEVGRVYGVTRIRQIESRTWAKL  
QHPARSGLGYLD"  
complement(6501..9518)  
/gene="asm30"  
complement(6501..9518)  
/gene="asm30"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450"  
/protein\_id="AM54108.1"  
/db\_xref="GI:21449376"  
/translation="WVATGTRIPKPLVGNLLDVLTSDDLTVDVFLDRCHREHG  
IVALTAGQKQVASSHELVARMCSPSGKAVHPALEQVRDPAGDGLFTARGDEPNW  
GKAHLLMPAGFPAMDIDAEQMLVRRPFGDHRIDVADDMTLTLDTIA  
LCAGFARFNSFYDRAPHPFDVAMVRLVEAGRAELRQVQPLVGRNQYRDDIATM  
NRADGVAARALPAGERDDLLERMLTCADPTVGERLSARNVYQLATFLTAGHET  
TSGLLFAVRLIAHPPELVKAKADVGLGDRVPAPEDLARDLYLQVLRLETLRHP  
TAPALAPPEALPGLGHAIGAGEPVLMLPTLHRDPVAVRDPDVPDFPERFAPRMD  
IPACAMFPGHGRACIGRFPALQEAATLVLLVLRQDLALADPDHRLTIKQTLTKP  
DSLVVRARPRDAGATATVETVPHQVPAIHRHGTPHPLVYFSGSGSGELARTIAG  
DGAARWATSVLDDRALPASGPVIVSSSYNGAPPDPAAHFVRNLQDGPDLISG  
VDYLVLGCLNDMSATYQVPTLIDEMAAGARLRERGAATDARADFFGDWRWIEP  
LWPLLSAEQVEIGEPFRVSVSDAAGDLGDLASAVLENRELVRGPTDGSKRHLE  
LRPDGTSYGTGDLVLPQNPDLVRAVARLCTRAERVVTVSSATGTLVPVGRAL  
RVDELLTRCDLSAPAGGVVARLAERCPPEAEIAATYGTLLERLERRPSCAVD  
LALALELLPRTLYLSISAAEQAEEVALTVSTVGTSGYLSVRPGRVAVGIAV  
PPESFPDPTVPLVLTAGTGIAPFRGLRARAALGEGPCPALLLFGCGPBLDDL  
YAEFFAALGWLVDRAYSRHPDGEVRHQRLRDRVRELVDAGARVYLCGDATR  
VGPAVEVLGRIGPGAGWLDALRAGRYATDVF"  
9664..10245  
/gene="asm29"  
9664..10245  
/gene="asm29"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="AM54107.1"  
/db\_xref="GI:21449375"  
/translation="MSAVRRQLVAAALVWRKRGICAAATTAICAEADMPHGAFHY  
CFRSQELYTALLTTLGNTSLDAWPAISPDLPVNDVTKMLAAFTWPAQADPERELV  
LGLVTLQARPELPELPAWQHREYVAKAVAHLDLBAEAGPDYSVDRLALBLVVS  
LDGVASSWFSDGELAKASLDQFAVMEALTR"  
10393..10857  
/gene="asm28"  
10393..10857  
/gene="asm28"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AM54106.1"  
/db\_xref="GI:21449374"  
/translation="MTDITTRHTFVQNAALAVGAVFLVGIAGFVPGLTNTVDGLEF  
AGHSDAALLGVFKVSALNVLHIAFGAAGIAAGLSKAPRGAFLIGGAIYAVLWL  
YGVVDEHSTANFVPLNAADNMLHLGLAVGMIGLGAALGRNAHYRAARSDD"  
complement(10912..11937)  
/gene="asm27"  
complement(10912..11937)  
/gene="asm27"  
/codon\_start=1  
/transl\_table=11  
/product="glucose 1-dehydrogenase"  
/protein\_id="AM54105.1"  
/db\_xref="GI:21449373"  
/translation="MPRLAEQVVVVVVGASSGIGRVTVLAFPAAGARVVCARNTRAL  
DGLVEVRGAGGRAVTAIDAEAAVAVADVVERFGRVDTWVNAAGTVYGRVED

TPAGEFDVMEVNYLGHVGHAKAALPALRRAGGVLLIGVASVGLSRLQAPQAYAAK

Query Match 5.2%; Score 26; DB 1; Length 82746;  
Best Local Similarity 100.0%; Pred. No. 0.088;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 ATCGACGCTACGGCAGCAGCAGCAGAC 446  
|||||  
DB 36138 ATCGACGCTACGGCAGCAGCAGCAGAC 36113  
|||||

RESULT 6  
AX417445 88421 bp DNA linear PAT 18-JUN-2002  
LOCUS Sequence 1 from Patent WO0231155.  
DEFINITION AX417445  
ACCESSION AX417445  
VERSION AX417445.1 GI:21522731  
KEYWORDS  
SOURCE Actinoplanes sp.  
ORGANISM Actinoplanes sp.  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Actinoplanes.

REFERENCE 1  
AUTHORS Zazopoulos E., Farnet, C.M. and Staffa, A.  
TITLE Gene cluster for ramoplanin biosynthesis  
JOURNAL Patent: WO 0231155-A 1 18-APR-2002;  
ECOPTA BIOSCIENCES INC (CA)  
FEATURES  
1..88421  
Location/Qualifiers  
source  
/organism="Actinoplanes sp."  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1871"  
misc\_feature  
/note="ORF 1; positive strandedness"  
misc\_feature  
/note="ORF 2; positive strandedness"  
misc\_feature  
/note="ORF 3; positive strandedness"  
misc\_feature  
/note="ORF 4; negative strandedness"  
misc\_feature  
/note="ORF 5; negative strandedness"  
misc\_feature  
/note="ORF 6; negative strandedness"  
misc\_feature  
/note="ORF 7; positive strandedness"  
misc\_feature  
/note="ORF 8; negative strandedness"  
misc\_feature  
/note="ORF 9; negative strandedness"  
misc\_feature  
/note="ORF 10; negative strandedness"  
misc\_feature  
/note="ORF 11; positive strandedness"  
misc\_feature  
/note="ORF 12; positive strandedness"  
misc\_feature  
/note="ORF 13; positive strandedness"  
misc\_feature  
/note="ORF 14; positive strandedness"  
misc\_feature  
/note="ORF 15; positive strandedness"  
misc\_feature  
/note="ORF 16; positive strandedness"  
misc\_feature  
/note="ORF 17; positive strandedness"  
misc\_feature  
/note="ORF 18; positive strandedness"  
misc\_feature  
/note="ORF 19; positive strandedness"  
misc\_feature  
/note="ORF 20; negative strandedness"  
misc\_feature  
complement(73563..74216)

```
misc_feature /note="ORF 21; negative strandedness"
              complement(74213..75424)
misc_feature /note="ORF 22; negative strandedness"
              75335..76464
misc_feature /note="ORF 23; positive strandedness"
              complement(76449..78110)
misc_feature /note="ORF 24; negative strandedness"
              complement(78107..79864)
misc_feature /note="ORF 25; negative strandedness"
              complement(79861..81624)
misc_feature /note="ORF 26; negative strandedness"
              complement(81682..81909)
misc_feature /note="ORF 27; negative strandedness"
              complement(82062..82346)
misc_feature /note="ORF 28; negative strandedness"
              82587..84446
misc_feature /note="ORF 29; positive strandedness"
              84481..85548
misc_feature /note="ORF 30; positive strandedness"
              85556..86845
misc_feature /note="ORF 31; positive strandedness"
              complement(86803..87372)
misc_feature /note="ORF 32; positive strandedness"
              87494..88420
misc_feature /note="ORF 33; positive strandedness; N-terminus only"
```

## ORIGIN

```
Query Match      5.2%; Score 26; DB 6; Length 88421;
Best Local Similarity 100.0%; Pred. No. 0.087;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 20 CGACCGGGCGGCCCAAGGCGTCGTG 45
    |||||
Db 38404 CGACCGGGCGGCCCAAGGCGTCGTG 38429
```

```
RESULT 7
AF007101 AF007101 32870 bp DNA linear BCT 13-JAN-1998
LOCUS Streptomyces hygroscopicus putative pteridine-dependent
DEFINITION dihydrogenase, PKS modules 1,2,3 and 4, and putative regulatory
            protein genes, complete cds and putative hydroxylase gene, partial
            cds.
```

```
ACCESSION AF007101
VERSION AF007101.1 GI:2624946
```

```
KEYWORDS Streptomyces hygroscopicus
SOURCE Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
        Streptomycineae; Streptomycetaceae; Streptomyces.
        1 (bases 1 to 32870)
```

```
REFERENCE Ruan,X., Staasi,D., Lax,S.A. and Katz,L.
AUTHORS A second type-I PKS gene cluster isolated from Streptomyces
TITLE hygroscopicus ATCC 29253, a rapamycin-producing strain
JOURNAL Gene 203 (1), 1-9 (1997)
MEDLINE 98085969
PUBMED 9426000
```

```
REFERENCE 2 (bases 1 to 32870)
AUTHORS Ruan,X. and Lax,S.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-1997) 47PAP9A, Abbott Laboratories, 100 Abbott
        Park Road, Abbott Park, IL 60064, USA
```

```
FEATURES Location/Qualifiers
            1..32870
```

```
source /organism="Streptomyces hygroscopicus"
        /mol_type="genomic DNA"
        /strain="ATCC 29253"
        /db_xref="taxon:1912"
        complement(108..1130)
        /note="orf5; similar to RapK"
        /codon_start=1
        /transl_table=1
        /product="putative pteridine-dependent dihydrogenase"
```

## CDS

```
1454..9625
/note="orf1; loading domain"
/codon_start=1
/transl_table=11
/product="PKS module 1"
/protein_id="AAC38061.1"
/db_xref="GI:2624948"
/transl_table="MLRRLIRPLPELLKHAARRGSQTAFADRRRSVSYAQLAARTR
RLAGHLAQLRLLPGDRAAIYLGNCVETIESYLAIRASAGVFPNPHSSIAELAYLLD
DSAARVITDHTLRPQVRLOAERPHLTVVVTGDESGAVFDALAEETPOOPARDLL
GLDDIAWMLYTSGTGPKGLSTQSRCLMSVAASAGVLGSETDRLVWVPLFHSLL
AHYCVLSTAVGATARI TEAFABELLHLRAEEFTFLAGVTWYLLDAARDGDL
SLPNRLVCLSGAISTATLRAEFOETRGVLLDCYSTETCGLITANWPNCTQAGSC
GLPVGLTVRIIDETHESVETEGEGLWVGPSLWYGNQPEATRAALPRGVRTG
DLGRDGLGHTITIRUSLEIIRSGENIHTEVEQVLLRVFVGTDAAVVGSAAHTLGE
VPVAFPAADGDFDPEFAACRHELAYFKVPEELYEIGSPRTGSGIKKHAQAQSP
ARLEAVSGSFPESLIRLEWTAPEAGFVDGQATSAHRAGSGTAGLEAPGSAVDC
PRGSLDEVLRAVRTGLTDRSGRVAVLTHGDHDPAAQAVMGLVESMADAPHDAPLV
DADGGEATAALSAVLSQGPALVRGCVLVRPLARAGSGQORVSRPDPDGTIVFGAD
QEQAELAGHLVAAYGIRHVLLARTAVPDLVIELTELGSVTCOTVRAGPRASLAE
FQTAGPPVAVLVQAGPELPTLTGVELGLQDALDELTAHPALVLSAPID
NPAPHAVVANNRHAAGPALASWGPREELTRERLALDFDAATVDDDAIRH
PQAGAAAGLAAATGLALPATIADFPNPDALAYLGGRLFGDTSAAGDFTPTPAAP
SGEPIATVGMACRLPGVTSPELWELLSSGGDGSIGFPTDRGWLDEHLVLPDHPG
TSYVREGFLHDAGDFAALGSI SPREALAMDPORLILTTWVLERAGIDPSLRG
SRGVFAGVMTHTYISRLRTPPELEGLVSGGSGSVSVRSVTFEGEFAVSDTIA
CSSLVALHLAQLRTECTLAVAGVTVNASPSYIEFGRGSLTDPGCKSPAAD
ADGAABEGATLLVRLSDAQRGLHGVVGTAVNQDASGLTAPNGPQORVI
RDALTAAGVATMDIDLTEAGTGTGLGDPTEAQLVAMYGQDRDLFVWGLSKNIGH
TOAAAGVAGVAKVLSMRHGMVSKLHIDEPTPEVDSAGTVELLAARDWPGADERP
RRAGVSFGSGTNAHVILQCAPAPAPPTESAEPSEAAAGPSGCGLELGPSPVP
WLSAQNETALKOLARQIVITGTEGTHQIAHSLITQSRSTTLPHRAVLGSTRREPATA
LEQTRHPDGGKTVLPFGGSGQWVGQRLMETSPLYRDTIHEIDTALSEYIDWSV
TDILTSQGAFTMDRVDIQVLETTMTALWQHLETPDAIIGHSGQIAAAYTA
GALTLOAAARTTLRSKALLPLTGHTWISVLAPPEKTOELADPHNWTIWAHNA
PNTTITGTOAI THLTQOLSHRIMRWQLPGVD FAGSHGIDQHLNQLKILAPITP
OKPTIPYSTDNTWITDILDADYWRNLQPVLFQPAINHLNTWNTVYIETSPH
TLTPIQETNPNTTITHLNNDTHALLTALGHATHGHPITWHTLIPKTTPLP
TYPFOHRYWLNDKPTVDQATQGLIPTPHPLVPAITTLAETGAILLGHINPAHH
PMTTHAVNGTPELPGTALVDMALHAGDHTDHTLDELIIHTPTLTHPTTITPIS
PDNTNTHLTIVHTRNSPDGPMWTAHATGLTOHQTPDPPTPTPTTDAEPIPLDGTQY
RLATGLYGTFOGLDARWRDLRDLVRLPNTDHTSDADGYGHPALDAAHAL
VAVSDGEIRLPEAWTGLHATGATKLRLVELEWDSASLSRAFPDQSPQVWVESL
ASRVISPEQLRSASARSGPLYRPWLIANSAGTGGAVFERVEYPTVGDVLDATHR
RSTETLRIQOHLAADITTLPLVBEATHDGLAAVWGLRTAQTEHFGRIIILITDNEH
PTSRKALPTIAGSEPOARDGATITPLRTVVAPEFAGVGEVGLITGTGSLGSL
IARHLVTEHGVRDLVLSRQGPAPGATELAEQLQAGVVRVSCDTIDRAALV
DEVGPFAGVHTAGALADATVDNLDALATTAAKAHAAWMLHEHTDGLITLTMFV
SSLASVLGSAGQANYAANSFLDALATHRTLGLPGTSLAWSWERTSGMSAHLTRAD
HDLRLARGLRPLTDEQGLEFPDGMALVDLQGVVATFDLPALGRSGVPLVLRVA
PRRAAGADSGGAEFWLTERIATLAPDARERSVLTSIEVQAQVALLDQVSPAPAAQFD
KSLGDSLTAVFEFRNLISAMTGLQLPATLVDFYPAAGVATRYLLDQVSPAPAAQFD
DPDERELRVLERVPAVFRFAGVLEALLKLAGDPDPARTPDAAEFDEEDGLIDMTEA
LIRHVMGTGA"
9637..14925
/note="orf2"
/codon_start=1
/transl_table=11
/product="PKS module 2"
/protein_id="AAC38062.1"
/db_xref="GI:2624949"
/transl_table="WTNEQLVEALRVSKENARLRRENASIQEGAAEPVAIVGMACRL
PGVDSPESLMELMASGDVSPFTDRGWLDDGLYDPPDHLGLTAYVREGFLDRAG
```

```

AFDPDPFGISPREALAMDPOORLLLETLWEVLERAGIDPTLLGRSGRTGVTGIMYDRP
VTRLRDFPRLGNGSNAASVASCRTSVYTFEGEPVNSLDTACSSLSVALHVAVRP
ALRRLIARVAGGVYVMSVHPYVEFSRQGLAFDARCKSPDASADGTVLSEGVGLLIY
ERUSDALRGHEVIAVVRGAVNQDASNGLTFTGTSQQRVIRDAITDARLTVADV
LLEAHGTGRLGDEIAQALIAITYGQDRQPVWLGSKNIGHTQAAAAGVAGVKTVM
AMRGVMPRTLVSEPTFNVDWSAGVRLTWARDFSTQDRERRAASVFGISGTNA
HVLEQAPAPEDPVPVTPAGVPWVLGAHTGPALSALAEALAAHVEDDEAPLGDA
HTLNTFVATLPHRAVVGTEBEFVAALRAVRPTRRGRGLPVMFTGQSQRPMGRQO
LVEHFPFAQALDEVALATPGIRVEMFDPDQAEITQRTDHAQIALPASETALRYLE
SWGLRDMVCGHSGEITAHVSGTTLTDDAHLVITGTLMOQLPFGGAMLAVAIDP
HTQLPHLDNHHDITISIAANGPHATVLSGDRITLHIAITQNTKTNWLVNSHAFSPFL
MQLPQFTTLNLTHPHPTPLISMLTATPTHTPTHTTQNTQITPAVRYTDTLHLH
HGHITVLEGPDTLITATLPTLTHLIPTRRNHNEVSTGTRDPLGLAVGVTSADV
DMRALPTGRTSLPTPFORRDFWLHDAAGGABEGAGLGITNDHPLGLAVGVTSADV
GELISLGRSLSTGPHMLDHTVNGTVIPGTALIDLALHAEDHTVDELVIHTPL
ALHTPSPQSTGAEDTGHPIALASRDITGTWTRHTGTLSNQTHPAAELTNSPT
DAHQIDTVAQQLADGLHYGPAQGLHTLYQORDNTLAFABIELPFAAGLPTGHILHP
ALIDALHPAAVAARDDESGIQLPFSWAGVTVHSTGATHRLVRLDITDGTARRLAT
DITGQVITIDSLTTRPLHPDQJHAAQEPNTNLYTLGWSPTPTPTTGHFERVYVPT
AGDILADTHRIITAOGLTRIQOHLAADTTPLVVEATHDDLACARVWGLRTAQTEHP
GRILIDTNHPTSRDALASVAGKQEARIDGAILPRELTSVLNTRASTRERDIL
SMGFLVGTGGGSLGSLIARHIVTEHGLDLVLVSRQGPADGATLNOLOQHGARI
RLTSCUDTDLATLIDELIGLTVVHTAGALADTTIDHNLPTTLITLTPKANPAW
WLHELTQDHLALFVSVSSVAGVGNAGQANYAANSFLDALLTHRRROGLPQSLAW
GMWRQEGTFAHITQADHDMTRAGIHGLTDAEGTTLFDLDRGLGAVPVLKHLPT
LNRADTVPAVLRLGLVRAVPAARQATVPAESWADRIAALMPEERDRAVLDRVQOQAR
VLGSHDADRIDPGCAFQDGTGFSLTAVEFRNRVTATGLRLPATLIFDHPPTNALADH
LLAQGLPLGDSGSSPEVVDLBERFERQLFDRPDDVEGTALIAVRLSLIHLDAFAT
APAAATASQVLDLTEDASOSSELLSFIDKEFGASN"
14964..21266
/note="orf3"
/codon_start=1
/transl_table=11
/product="PKS module 3"
/protein_id="AAC38063.1"
/db_xref="GI:2624950"
/translaction="MADSKTLDYLKWVTADLYTRERLAEVESAEPATIAIGMACR
FPGGVTSPEALWELVSGTDAISGFDDRCMDWEGFPDPAQBGTSYMGEGFLTNA
YTGDPDAFGISPREAVAMDPOORLLLETLWEVLERAGIDPTLSRGSTGVTGTMANS
YTRVRRIIPGELISYAGNAGSASVASCRTSVYTFEGEPVNSLDTACSSLSVALHVAVR
ALRSRGCEIAVAGGVVATPRIPLFSLRQRLGARDGRCAFSADTGLSGVGLL
LVLRLSDAGTGLSVLGVGTAVNQDASNGLTFTAGNPAQQRVIAKALADARLRTKI
IDILEAGTGTGLGDEIAQALIAITYGQDRQPVWLGSKNIGHTQAAAAGVAGIIKI
NAMHVLQAPAPEDPVPVTPAGVPWVLGAHTGPALSALAEALAAHVEDDEAPLGDA
VAMRHGIMWPTLHIDPEVDTWVTEVTEVTEVTEVTEVTEVTEVTEVTEVTEVTEV
LAEQLATVDSVADIAHSLISTAAALPHRAVVGSTREEFATALEQITGHTPGGKTV
LVTFPGQSGVWQGRUWETSPLYRDTIHEIDTALSIEDVSDITLDTLOFGAPTWR
VVIQVPLFTTMTALATLWQHGLTDAIIGHSGQETAAATAGALTQDAPRITTLR
SKALLPLTGHGTWISVLAPPQTIQDIADPDHWNTRIWAIAHNAAPNTTITGDTQAITH
LTQOLGHRIMRWQLPGVDFAGSHIDQLNQLKILAPITPKQPTIPYFTDNTW
ITPTLTDADVYRNLRQVPLFOALNHLTNNGVITYLETSPLTLPYSIOETNPNTT
IHTLNSQNDTHALLTQALGHANTGHPITWETLIPPTKPTPLPTYPQHTPWLNETT
ADADVSAGLSURDTPULGLATLHSGETVLTHSGTTHHGMADHAVNGTLPPLPT
ALVDMALHAGDHTDNLDELIIHTPTLTHPTTIQUTTTTHDDNTLNTLITRYTAP
EQPMIOHAQATLRANTAPSPETIPPTNAEPILPDGTQYRLATGLLEYTGFAPQGLH
ALVRNSDNLHATDVRPDTIDNTNSDCTDGYCIHPALDAAHALVAVDGSDELR
PEAWTGVTHATGATKLRVLDIDDDGTLSDLSATDSQPVVTVDSITTRPIDPAQLR
ARPRQDLDYALOWLPNSGVSAPAPEQYEVKVGDDVLADTHRTAQTLTVQOHL
AADTPTLVEATHDHLAAVWGLIKTAQTEHPRILIDITDNHPTSRDALASVVAS
GEQFQARLDGAILTTLTAVTADTALTPPADADAHLDTSRGLENLTITSGDTSATR
NLLPGEVRAVRAAGNFRVLIALGMLNGEATMGSEAGVILVDIGSDITDLPGRV
TGLFPNALANTAVAGNRMITPIPASWNSNTDAAAPPIVLTATYGLHDLAALKPEGRV
IHAGTGCVGTAAITQALHCAEVPATAPTKWPHLRAMGLDHDHIAISRTPDPTTFH
TTTQCGQIHLVLSMVDLTDASLRLITPNGRFTEMKTDVTRTEEVATHRGVTRYRA
FDIADAGPRGIMVMDLLETLGFERGALRGLPVRCDVRESVDAFRFMAQARHIGKV
LATIPPTDGLTITGGTNGLSLHIAHLVTHGHARDLITLSRQGNAPAGTLELNTQIQ
QHGARIPTISCDLTRTALATLIDEIGLTVGVTAGALADTTIDHNLPTLPTLTPP
KANPAMWHLTDHDLALFVPSVAGVGNAGQANYAANSFLDALLTHRRROGLP
GTSIAGWQMOREGCTAHLTQADHDMTRAGIHGLTDAEGTTLFDLDRGLGAVP
KLHLPTTARGSVLPVLRALVRAVPTAGQAQVODGTWADRAVAAPERRAVVLELV
RAQATTVLHGSHGAEHVSPTQAFKDSGFLSTAVEFRNRVTATGLRLPATLIFDHPPT
NTADHLLAQALPATYGTGFSLTDGIEVTLRTLAEDGERRPAVVVRELKQWLSTIGAAQ
PAADTALDLDAASARETFDPLDGGAG"

```

```

21363..27176
/note="orf4"
/codon_start=1
/transl_table=11
/product="PKS module 4"
/protein_id="AAC38064.1"
/db_xref="GI:2624951"
/translaction="MANEDFHLKKAADLRDARQIQLEAREPREPIAIVSMACR
LPGVDPSPLMDLVAAGDGVSAFPIDRGMDLEGIYDPPDHPGTTYARDGGFLDDA
AGDPDAFGISPREALAMDPOORLLLETLWEVLERAGIDPAAIRGSRGTGVTGIEND
YAWSADDTQLEGLYLGNSAASVASCRTSVYTFEGEPVNSLDTACSSLSVALHVAVR
ALRSRGCEIAVAGGVVATPRIPLFSLRQRLGARDGRCAFSADTGLSGVGLL
LVLRLSDAGTGLGDEIAQALIAITYGQDRQPVWLGSKNIGHTQAAAAGVAGIIKT
IDILEAGTGTGLGDEIAQALIAITYGQDRQPVWLGSKNIGHTQAAAAGVAGIIKT
VAMRHGIMWPTLHIDPEVDTWVTEVTEVTEVTEVTEVTEVTEVTEVTEVTEV
NAHVLEQAPAPEDPVPVTPAGVPWVLGAHTGPALSALAEALAAHVEDDEAPLGDA
SLHGRKAARDFCAVMIGSTREDFAELALTCGSRASKRVVFPFGQSGMWIWMGRGL
MESSAVRESVLRTDRALAEFVDMASAAVLAGEPTPDLDRVDVQVPLFTVMVALAN
VWRSYGVBPAAVVGHSQGETAAAHVAGGLSLRDAARVVALRSALATALAGEGVMVYQ

Query Match 4.8%; Score 24; DB 1; Length 32870;
Best Local Similarity 100.0%; Pred. No. 0.92; BCT 15-MAY-2001
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 429 CTACGGCAGCACCGACGACCTGCGG 452
|||||
Db 2383 CTACGGCAGCACCGACGACCTGCGG 2406

RESULT 8
AP002998
LOCUS 347750 bp DNA linear BCT 15-MAY-2001
DEFINITION Mesorhizobium loti DNA, complete genome, section 5/21.
ACCESSION AP002998 BA000012
VERSION AP002998.2 GI:14022369
KEYWORDS
SOURCE Mesorhizobium loti
ORGANISM Mesorhizobium loti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE
1 (sites)
Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,
Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,
Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M.,
Takeuchi, C., Yamada, M. and Tabata, S.
Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)
21082930
PUBMED 11214968
REFERENCE 2 (bases 1 to 347750)
Kaneko, T.
Direct Submission
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp,
URL:http://www.kazusa.or.jp/rhizobase/,
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
On May 11, 2001 this sequence version replaced gi:11994973.
FEATURES
Location/Qualifiers
source 1..347750
/organism="Mesorhizobium loti"
/mol_type="genomic DNA"
/strain="WAF303099"
/db_xref="taxon:381"
complement (12..2333)
/gene="mll1656"
complement (12..2333)
/gene="mll1656"
/codon_start=1
/transl_table=11

```

```
gene
/product="ATP-binding protein of dipeptide ABC
transporter"
/protein_id="BAB48979.1"
/db_xref="GI:14022370"
/translation="MNEAVRNPAPEIIEIENLSISFFTRKGEIPAVMDFSCTVMPGEA
MGIIVGSGCGSTVSLGIMRLDSNIKTVGRIRKFGQKDMGELSEELRAIRGNKIAM
IYQPMASLNPAMKVGQOLMEVPLIHDKVSKEEAYKRSLEVMRAVKLPDPERMWSYP
HOLSGQOORLVIAWALLSKPALLDEPTTALDVTVEAGIVELVKGIGKFGTSMIF
VSHNLGLILETCDRTIWMYSEAVETKIKDVFDRMRHPYTGQLFRSIPLPADKNSR
PLISIPGQPLPFRKPGKNGPRCHHEFVEGVCNAAEIPMVEVAGHGFGRKCRKFN
IDWALPFGAKVNERVVPVGLKIEDLRYIKVGGSEVFGSGSEGRVYKANETISSI
ARESETVAIVGESGCGSTLAKVLIGLETASAGTVLGNKOIQSGIESRSETVSSI
QMVFNPDPNPNHSGVSOIIRLTLEKNVGTADVDRKRMLELIDVLKLPRAETFK
PROLGGOKORIGVARAFAGAKVADPEVSALDVSQAAVETLMDIOKRNKTMIL
FISHDLSVRIARUVVMYLYGYVEQGTQIQIEAPPHPYTEALLSALPIADTSYVK
RHVLEGGIDPSAMPNPCTPCTRCGYKKIKVDNLCETKPPVVKHLDGHSCLWAD
DLVAKMEFVIRKDEHAAHEGVDPDAPHTGTFGEFASPPARPSPGKNAAEADATGKAV
EDTDAVLRARHEVRDETSERGSKPKDTRRH"
complement(2382..3539)
/gene="ml11657"
complement(2382..3539)
/gene="ml11657"
/codon_start=1
/transl_table=11
/product="permease protein of dipeptide ABC transporter"
/protein_id="BAB48980.1"
/db_xref="GI:14022371"
/translation="MQVQIISAFITVILEVSRFPWPWTALIIVMGASFYKKKLALYG
QLFDSGVIGVIGICLFWLFTAFIFASTISPFDPDLAQVPMKMDLPGLTEPOSGLYLF
GLDKLARVDFRSMVSGQITLIIAPAATGALMVGITLGLPAGYTGKIDTIVLSFAN
LVPLATLIFLLYTVGTMIDPIPYAMAGVFPFPIIFFSVFWTRKYNRPDRILYL
LGLTLIGCVLIGESVFWDRDPLHVIDPNQLNIKVAVFASSPGVRIVRGLWMDIK
TRDYVAAQTRGESFWYIMWELLNARGPLIVDACRLGYTTLIGLTLGVLGLGLAP
ESPMDTAKDRASRLRSLFIHPALPPTIALMSFVLGRLNLLDSLRQSKMD"
complement(3552..4586)
/gene="ml11659"
complement(3552..4586)
/gene="ml11659"
/codon_start=1
/transl_table=11
/product="permease protein of oligopeptide ABC
transporter"
/protein_id="BAB48981.1"
/db_xref="GI:14022372"
/translation="MLSFVIRRLGTWALTMCLTMTWVPLINLEPNLKLALISQEMH
TSAEQLESLWNHGRQNFVRYQWLGVMKPQVTDPAIGKPAQRFSCNDPVPVTF
AGVVEGDFGCSFTKFTVASKLFPALGATGILMFCVLAWMVPVSLIIGILAGMRGSR
TDRVLSVASIATPEYVSGVIFTVIFASWLGWNGSASASQGITFYNFPLPVMTL
ATYGIGYIARMTASRMVENVTOQYIIRTLKGLSGFVGVVVKHALNALIAPFTVIMLQ
FPMLLTGTVIVVEVMFYQGFYTLVEAAGNNDIDLLGCSLYSVFVILTITQLSDVGY
AFLNPRIRVQ"
complement(4809..6464)
/gene="ml11660"
complement(4809..6464)
/gene="ml11660"
/codon_start=1
/transl_table=11
/product="dipeptide binding protein component of ABC
transporter"
/protein_id="BAB48982.1"
/db_xref="GI:14022373"
/translation="MTIYKSGNDVRPDHILMAEEHAKGVDRERFLALASVFGASTA
MAYGLLGLAPTEAREADAPKGGILKVAQWIKDPDKPSDSEIANNEROALEPLIV
KTYTETFRPYLLESWVDNDATQYVLVHRKGVTVHNGDAFTADDVWVHFNWGEKAA
EGNSMPRLGTLTFDDTKLREGAVTKDDMTVKLSLPEDIISIIPGICDYVALIVHK
SFDEKGGDFKACPIGTGFPELSDYDVGVKVRADNKWMDGVYLDGIEDTGYD
PAAWVSAPFAGEVHTFTSADYVSILDKMDLVKRSWNTASTVCTVNTVTHKPDNEK
VRKALQGVNNVLOLGYGNAGSVAENHVCISLHPYYELPKLTPDPAKALMAEA
GQADFHELITVDVEDWHKNTGDIAAQLRDAGIKVRKTVLPGSTFWNDWIKPLSWIN
WNNRPLGVQVLYATXETGEAWNECAWSNPNDPAKNAALASDAAKRKMMDIQIL
QDSGIITQPYWRKLYSHKVAQVNYKHPTFFEDYDGIWLDQA"
complement(6717..12953)
/gene="ml11661"
```

```
complement(6717..12953)
/gene="ml11661"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48983.1"
/db_xref="GI:14022374"
/translation="MPVRCRRSFPTSPTRSAPASACATRRRSACRSMRPFSTAI
PTIRILASPASGRSRDMKTVMRILRIVLVTLIVAGALVLTTERGRDLNLAG
IISRLASDDRRKYTVSGDGIWSGALSVDHVLEDRQGAWLVARVWDSPFALLSK
SFSADRIAAREIELARLEPVVGTQPPAQGGTITLIPVSDIREIDLEPEIALGQOLASGI
AELAKMLKADAAPLAVETVNTVRHDKGQNDVAKIHFADPADRLDLGKSPAG
GIILANLLKPTPPVDVAVSGTGPLANNGIATFVVGKIVTQLTGRQLTGHKNHVE
AKGGEPAFPDNLRSFAGKTGFDAWGATAGGCSIDRASIESDAIHGTATGIAD
PAGASDLVETAAKSAPVQIVVGSAAQPIITVAIRNATARAFDQKAPITVDIGASLASV
VAGTQNLNIDSAQIHSQDFDQNRAGPVAVKLTAGGLTDVATLAPLTVKGVDDLAG
SLSDATAIDQSDNLNALSMTVALDSLSASGTASAGSDIQADIKGTIGDVSLSVMV
VKFSAAATRPQGAFAANISILTSGLSASGTASAGSDIKIQTASGTINPGASFDLDT
GVPVAGVNFALSGAGARTAPDFTVSADSDLSITAGRTVXAIKLTASGADITASPSAD
VSLTGSVNDQPLDLKAALVTSEGRKRSINGFLVSLGDNKVSGLDLDKDFMPLGTLTAL
APADOLAALAGQAVTGDIINGTILFAKGDAPSVTIDAKTSIARGEVTAIAVNAL
IANYLKAPASISGTVKADSVTSGATEISIGIVDLKEDGDMTNGTGTAGIAPATAAGR
VKIAQGTTSVEIASGEATVRGKAAIAEPSTIITANGTTSIEKLLDVGGSVTLISGS
AQOTLDLAAQFSAALPAALANDFAPGLDAVTLGGAAQVTSGPSAEPVFNQAQGLVET
SQTRQAEGLPLAVDAAGVYITLAGVVALDHATLSGDKISGTASGTINPGASFDLDT
SSGSLPLTVTGSAAAPVRIEIRSLSAKVAGESTRARLDVSAIPLPITTSIPAKVDGLVL
ALHSDAPLKNRAGPVSGTVSDVKVLDNAIAPLIAGKVTAAANARLTSQGVADISG
SLKSDANQVAGVSRLDGDALDLKLADAPSSALPAARGMLGQAOISATLKRDPN
GGINVDGLVSGGLTAQOASLADNKLADITGALADVSLSKDAGAIAPALNAQG
PAIAPDLSLTVNSRLLVAEREITGLKLTATGKADAANPAANVQUTGVAGALOGSA
VLATADGSAINGLLLSLGRKISGLDLDKDFVEPTISLDPLDIGLALAEKAE
GDRGTIAPTNGDAAPVAIKAAATASITRGDLQAKTVSIDALIANVLAAPVTSKIRA
DVTVSGGTVISGIDVLDKRDGDTGSGSATVKGIPAKAAGRVKAVANGTTTIELASQO
ATVQIKAAIAQAASVTSIANGTTTLDRLVNLGGTATVTVGKQALDINAVLKVPM
SLANGFSPGLDAAGSISGTVKVTGAPSPAPAFNIDAGVQTSQTRGAGVAGVSSSS
GTAFGNKLTFNANIDGAGLGLGGTVTITAGTPALVDNFVAGVPMUSKLAAGGL
SLSGTANVNVQVRGSPATSGVSGSVTSGLARLDARSLVADNLAADVSIGGVARIN
RLTGTLSRGLSASGTGVINPAQGFADLSIKLTDGRVTDGRVTANLGGDLITKGP
LVSPVIAGTYNLAKTVITVPKLPASISALVDKHNAPEAQAQKALRPATTSKGS
GSGSLADVTYNAPNQIIFIQGRGVDAELGGSLRLTGPASSPOAVGTFTLQGRSLTILG
KRLTFTGTVGFGSGLPYLNLATTTTTSATVITVVSGEATNPKFTFSFSPALPEDE
VLAQILFGRSMNSLPLQIAQLAEAGQLAGVGGSTSLLENLSAIGVDDLDVTTDDK
GGTAVSAGKYLNDRTYVTITQDKPFGSKATIDLVNVRGVKLRGEATDAGEARGGIFY
EKEY"
```

gene

CDS

gene

CDS

gene







```

/contig_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10984.1"
/db_xref="RENTREMBL:CAB10984"
complement(1542..1682)
/note="MLCB1779.04c, probable isocitrate dehydrogenase
(nadp) pseudogene fragment, len: 46 aa; similar to part of
many eg. IDH SPHYA P50215 isocitrate dehydrogenase (nadp)
(406 aa) (56.5% identity in 46 aa overlap)"
/pseudo

/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10985.1"
/db_xref="RENTREMBL:CAB10985"
complement(1684..1941)
/note="MLCB1779.05c, probable isocitrate dehydrogenase
(nadp) pseudogene fragment, len: 85 aa; similar to part of
many eg. IDH SPHYA P50215 isocitrate dehydrogenase (nadp)
(406 aa) (63.6% identity in 33 aa overlap)"
/pseudo

/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10986.1"
/db_xref="RENTREMBL:CAB10986"
complement(1990..2106)
/note="MLCB1779.06c, probable isocitrate dehydrogenase
(nadp) pseudogene fragment, len: 38 aa; similar to part of
many eg. IDH SPHYA P50215 isocitrate dehydrogenase (nadp)
(406 aa) (61.1% identity in 36 aa overlap)"
/pseudo

/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10987.1"
/db_xref="RENTREMBL:CAB10987"
complement(2107..2295)
/note="MLCB1779.07c, probable isocitrate dehydrogenase
(nadp) pseudogene fragment, len: 62 aa; similar to part of
many eg. IDH SPHYA P50215 isocitrate dehydrogenase (nadp)
(406 aa) (56.5% identity in 62 aa overlap)"
/pseudo

/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10988.1"
/db_xref="RENTREMBL:CAB10988"
2820..3149
/note="MLCB1779.08, probable pseudogene fragment, len: 109
aa; similar to part of MT17 YEAST P06106
o-acetylhomoserine sulphydrylase (443 aa) (33.7% identity
in 89 aa overlap)"
/pseudo

/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10989.1"
/db_xref="RENTREMBL:CAB10989"
3083
/note="conflict: this base is missing in EMBL:ML022 M.
leprae cosmid L308"
3208..3210
/note="conflict: ACG is GC in EMBL:ML022 M. leprae cosmid
L308"
3242
/note="conflict: this base is missing in EMBL:ML022 M.
leprae cosmid L308"
3429..3782
/note="MLCB1779.09, probable pseudogene fragment, len: 141
aa; similar to part of MT17 YEAST P06106
o-acetylhomoserine sulphydrylase (443 aa) (37.7% identity
in 130 aa overlap)"
/pseudo

/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10990.1"
/db_xref="RENTREMBL:CAB10990"
3767..4096
/note="MLCB1779.10, probable pseudogene fragment, len: 109
aa; similar to part of MT17 YEAST P06106
o-acetylhomoserine sulphydrylase (443 aa) (36.0% identity
in 111 aa overlap)"
/pseudo

Query Match 4.6%; Score 23; DB 1; Length 43254;
Best Local Similarity 100.0%; Pred. NO. 2.7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CDS
208 GCCGACGGCAGCTCCGCCGACGA 230
|||||
21891 GCCGACGGCAGCTCCGCCGACGA 21913

RESULT 10
MLEPRTN3/c 312050 bp DNA linear BCT 20-FEB-2001
LOCUS Mycobacterium leprae strain TN complete genome; segment 3/10.
DEFINITION AL583919 AL450380
ACCESSION AL583919.1 GI:13092766
VERSION
KEYWORDS
SOURCE Mycobacterium leprae
ORGANISM Mycobacterium leprae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium.
REFERENCE 1 (bases 1 to 312050)
AUTHORS Cole, S.T., Bigmeier, K., Parkhill, J., James, K.D., Thomson, N.R.,
Wheeler, P.R., Honore, N., Ganier, T., Churcher, C., Harris, D.,
Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R.,
Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C.,
Maclean, J., Moule, S., Murphy, L., Oliver, Quail, M.A.,
Rajandream, M.-A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S.,
Simmonds, M., Skelton, J., Squares, R., Stevens, K., Taylor, K.,
Whithead, S., Woodward, J.R. and Barrell, B.G.
TITLE Massive gene decay in the leprosy bacillus
JOURNAL Nature 409 (6823), 1007-1011 (2001)
MEDLINE 21128732
PUBMED 11234002
REFERENCE 2 (bases 1 to 312050)
AUTHORS Parkhill, J.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium
leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome
Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk

COMMENTS
Details of M. leprae sequencing at the Sanger Centre are available
from http://www.sanger.ac.uk/Projects/M_leprae/ A relational
database containing the M. leprae sequences is available from
http://genolist.pasteur.fr/Leprona/.
location/Qualifiers
1. 312050
/organism="Mycobacterium leprae"
/mol_type="genomic DNA"
/strain="TN"
/db_xref="taxon:1769"
complement(2412..3665)
/note="LEPRT, copy 1, 1254 bp. Shows 99-100% identity
with regions 72707..728460, 1539937..1541188,
1480312..1481018, 2203983..2204515, 691673..691731"
complement(2413..2473)
/note="61 bp repeat found at 3'-end of LEPRT and 5' to
REPUP"
complement(5611..6426)
/gene="ML0569"
complement(5611..6426)
/gene="ML0569"
/note="Similar to M. tuberculosis Rv1433, putative
membrane protein, TR:006825 (EMBL:AL123456) (271 aa);
Fasta score E(): 0, 68.3% identity in 271 aa overlap."

FEATURES
source
repeat_region
repeat_region
gene
CDS

```

Shows weak similarity to other hypothetical proteins in M. tuberculosis. Previously sequenced as TR:Q49706 (EMBL:U00013) (271 aa); Fasta score E(): 0, 100.0% identity in 271 aa overlap.

Similar to ML0426, ML2446 and ML2664

```

/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAC30077.1"
/db_xref="GI:13092768"
/db_xref="SPTREMBL:Q49706"
/transl_table="MGIMRAAIRGVLLVIGIAATVIAQFVRVSLAAVNHAGFAITSI
LPTDAVGVAPVTVTFCAPVTRKATELSLKVRSAFMNGKFDMLDSKVQWVPDR
YWPSTTIALTVGNLSNFTKGTAILADIASHFTFTVTDGVEAETPPPLSPHRRP
HWREGVWPASMGKTEFPPTGKTVMSKORSVIMDSVSGIPVDDPEGYRLSDVYAV
RISRGKLIHSAPWAVQSMGFNTSHGCLISLSPADAENYNTVNIQDPFVYKE"
6976..6882
/notes="possible RBS"
6889..7908
/gene="gap"
/notes="synonym: ML0570"
6889..7908
/gene="gap"
/EC number="1.2.1.12"
/notes="Similar to M. tuberculosis gap, Rv1436,
glyceraldehyde 3-phosphate dehydrogenase, SW:G3P MYCTU
(006822) (339 aa); Fasta score E(): 0, 89.1% identity in
339 aa overlap. Similar to many e.g. Mycobacterium avium
gap, glyceraldehyde 3-phosphate dehydrogenase,
SW:G3P MYCAV (P94915) (339 aa); Fasta score E(): 0, 87.9%
identity in 339 aa overlap. Previously sequenced as
SW:G3P MYCLE (P46713) (339 aa); Fasta score E(): 0, 99.7%
identity in 339 aa overlap. Contains Pfam match to entry
PF00044 gpdh, glyceraldehyde 3-phosphate dehydrogenases.
Contains PS0071 Glyceraldehyde 3-phosphate dehydrogenase
active site."
/codon_start=1
/transl_table=11
/product="glyceraldehyde 3-phosphate dehydrogenase"
/protein_id="CAC30078.1"
/db_xref="GI:13092768"
/db_xref="GOA:P46713"
/db_xref="SWISS-PROT:P46713"
/translations="MTVRVIGFGRIGRNFYRALLAQOEHGIADVQVVAINDITDNS
TLAVLLKFDLSILGRPHDVSLEEDTIVGSEKIKALAVREGPAALPWHAFGVGVVVE
STGLFTNAKAKHLEAGAKKVIYAPATDPDITIVGVNDDKVDGSONIISASCTT
NCLAPLAKVLHDQFGIVGLMTVTHAYTQDNQLQDPHSLRPARAALNVPTSTGA
AKAIGLWPELKGIDYALRVPIPTGVSVDLTDLSKVSVEINAVFQDAEGRUK
GILKYVDAPTIVSDIVTDPHSSIFDSGLTKVIASQAKVSWSDNMGWSNRLVDLVGL
VGKSL"
6892..7902
/gene="gap"
/notes="Pfam match to entry PF00044 gpdh, glyceraldehyde
3-phosphate dehydrogenases, score 733.80, E-value
7.9e-219"
7354..7377
/gene="gap"
/notes="PS00071 Glyceraldehyde 3-phosphate dehydrogenase
active site"
7911..9161
/gene="pgk"
/notes="synonym: ML0571"
7911..9161
/gene="pgk"
/EC number="2.7.2.3"
/notes="Similar to M. tuberculosis pgk, Rv1437,
phosphoglycerate kinase, SW:PGK MYCTU (006821) (412 aa);
Fasta score E(): 0, 80.6% identity in 412 aa overlap.
Similar to many e.g. Corynebacterium glutamicum pgk,
phosphoglycerate kinase, SW:PGK CORGL (Q01655) (403 aa);
Fasta score E(): 0, 59.2% identity in 407 aa overlap.
Previously sequenced as SW:PGK MYCLE (P46712) (416 aa);
Fasta score E(): 0, 99.8% identity in 416 aa overlap.

```

RBS

gene

CDS

misc\_feature

misc\_feature

gene

CDS

Query Match

Contains Pfam match to entry PF00162 PGK, Phosphoglycerate kinases. Contains PS00111 Phosphoglycerate kinase signature.

```

/codon_start=1
/transl_table=11
/product="phosphoglycerate kinase"
/protein_id="CAC30079.1"
/db_xref="GI:13092769"
/db_xref="GOA:P46712"
/db_xref="SWISS-PROT:P46712"
/transl_table="WRIPNLKDLLEEGVGRCLVRCNDLVFLGDDGAIITDLGRVTAS
VPTKALCEAGAKTVAAHGRKPNKPKLSLEPVAALGEQLQCNQVLVCSTDRSP
VGGDALACVERITDGDILLQNPREDPKTSKDDERIALAKOLVELVGSACAFVSDG
FGVHRRQASVTDVAITLPHHAGILVADERILEOLTISAPYAVLVGSKVSKLIG
VIESLATKADSVLGGGCMFTFLAAQGSFVSGSKLLETETFCRSLLITDYADVLRLPM
DIVTEKEVADSPQTVAAADAI PDGLMGIDIPGESVKRFATILSNASTITFWNGPVGVF
EFPAACTRGVAAEIVAVTGKAFSVVGGSDSAAARALSLPEGSVSHLSTGGGASL
EYLEGKTLPGIEVLGREQPRGDS"
7911..9140
/gene="pgk"
/notes="Pfam match to entry PF00162 PGK, Phosphoglycerate
kinases, score 757.00, E-value 7.9e-224"
7962..7994
/gene="pgk"
/notes="PS00111 Phosphoglycerate kinase signature"
9149..9153
/gene="pgk"
/notes="possible RBS"
9158..9943
/gene="tpi"
/notes="synonym: ML0572"
9158..9943
/gene="tpi"
/EC number="5.3.1.1"
/notes="Similar to M. tuberculosis tpi, Rv1438,
triosephosphate isomerase, SW:TPIS MYCTU (008408) (261
aa); Fasta score E(): 0, 83.9% identity in 261 aa overlap.
Similar to many e.g. Corynebacterium glutamicum tpi,
triosephosphate isomerase, SW:TPIS CORGL (P19583) (259
aa); Fasta score E(): 0, 64.5% identity in 256 aa overlap.
Previously sequenced as SW:TPIS MYCLE (P46711) (261 aa);
Fasta score E(): 0, 99.6% identity in 261 aa overlap.
Contains Pfam match to entry PF00121 TIM, Triosephosphate
active site."
/codon_start=1
/transl_table=11
/product="triosephosphate isomerase"
/protein_id="CAC30080.1"
/db_xref="GI:13092770"
/db_xref="GOA:P46711"
/db_xref="SWISS-PROT:P46711"
/translations="MSRKSLIAGNWKMNLMHFEALVQKIAFSLPKYYDKVDVTVL
PPTDLRSVQTLVDGDKRLTYGAODLSQHDIGATYDQISGAFKLGCSFVLVGHSE
RRYHDEGALVAAKTAALKNSLTPTVICIYEIVREYGVHSHQCNQLRGSILAGLS
PROIGNVIVIEFWALITGRVASADAQEVCEAIRKELGALASQVAVETVRVLYGGS
LNAKNTGDIVAQDDVDGGLVGGASLDGAQATLAVIAGGFLP"
9164..9931
/gene="tpi"
/notes="Pfam match to entry PF00121 TIM, Triosephosphate
isomerase, score 524.60, E-value 3.7e-169"
9665..9697
/gene="tpi"
/notes="PS00171 Triosephosphate isomerase active site"
9927..10202
/gene="ML0573"
9927..10202
/gene="ML0573"
/notes="Unknown function. Previously sequenced as TR:Q49684
(EMBL:U00013) (91 aa); Fasta score E(): 0, 98.9% identity
in 91 aa overlap. Contains Pfam match to entry PF00121

```

misc\_feature

misc\_feature

gene

CDS

Query Match

```

Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GCCGACGGCAGCTCCGCGACGA 230
|||||
Db 112474 GCCGACGGCAGCTCCGCGACGA 112452

RESULT 11
BD262946 7887 bp DNA linear PAT 17-JUL-2003
LOCUS
DEFINITION
  DNA sequences for enzymatic synthesis of polyketide or
  heteropolyketide compounds.
ACCESSION
  BD262946
VERSION
  BD262946.1 GI:33072714
KEYWORDS
  JP 2002527067-A/11.
SOURCE
  Polyangium cellulosum
ORGANISM
  Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
  Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
  1 (bases 1 to 7887)
AUTHORS
  Beyer, S., Bloeker, H., Brandt, P., Cino, P.M., Dougherty, B.A.,
  Goldberg, S.L., Hofle, G., Mueller, R.J. and Reichenbach, H.
TITLE
  DNA sequences for enzymatic synthesis of polyketide or
  heteropolyketide compounds
JOURNAL
  Patent: JP 2002527067-A 11 27-AUG-2002;
  BRISTOL MYERS SQUIBB CO
COMMENT
  OS Sorangium cellulosum
  FN JP 2002527067-A/11
  PD 27-AUG-2002
  PF 11-OCT-1999 JP 2000576029
  PR 09-OCT-1998 DE 198 46 493-2
  PI STEFAN BEYER, HELMUT BLOEKER, PETRA BRANDT, PAUL M CINO, BRIAN A
  PI DOUGHERTY,
  PI STEVEN L GOLDBERG, GERHARD HOFLE, ROLF JOACHIM MUELLER, HANS PI
  REICHENBACH
PC
  C12N15/09, C12N1/21, C12N9/00, C12P1/04, C12P7/26// (C12N1/21, C12R1: PC
  01),
  PC (C12N1/21, C12R1:19), (C12N1/21, C12R1:465), C12N15/00 CC DNA
  sequences for enzymatic synthesis of polyketide or CC
  heteropolyketide
CC compounds
FH Key Location/Qualifiers
FT source 1..7887
FT
  Location/Qualifiers
  1..7887
  /organism="Sorangium cellulosum".
  /organism="Polyangium cellulosum"
  /mol_type="genomic DNA"
  /db_xref="taxon:56"

ORIGIN
Query Match 4.4%; Score 22; DB 6; Length 7887;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 CGACCGGGCGGCCCAAGGGCGT 41
|||||
Db 1904 CGACCGGGCGGCCCAAGGGCGT 1925

RESULT 12
AF3211122 9858 bp DNA linear BCT 29-OCT-2002
LOCUS
DEFINITION
  Streptomyces antibioticus simocyclinone biosynthetic gene cluster,
  partial sequence.
ACCESSION
  AF3211122
VERSION
  AF3211122.1 GI:11321496
KEYWORDS
  Streptomyces antibioticus
SOURCE
  Streptomyces antibioticus
  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Streptomycineae; Streptomycetaceae; Streptomycetes.
1 (bases 1 to 9858)
Galm, U., Schimana, J., Fiedler, H.P., Schmidt, J., Li, S.M. and
Heide, L.
Cloning and analysis of the simocyclinone biosynthetic gene cluster
of Streptomyces antibioticus Tu 6040
Arch. Microbiol. 178 (2), 102-114 (2002)
22109063
12115055
2 (bases 1 to 9858)
Galm, U., Li, S.-M., Schimana, J., Fiedler, H.-P. and Heide, L.
Direct Submission
Submitted (13-NOV-2000) Pharmazeutische Biologie, Pharmazeutisches
Institut, Mikrobiologie/Biotechnologie, Universitaet Tuebingen, Auf
der Morgenstelle 8, Tuebingen 72076, Germany
Location/Qualifiers
1..9858
/organism="Streptomyces antibioticus"
/mol_type="genomic DNA"
/strain="Tu 6040"
/db_xref="taxon:1890"
100..810
/gene="simK"
100..810
/gene="simK"
/notes="putative reductase; similar to NovK of novobiocin
biosynthetic gene cluster; involved in aminocoumarin
biosynthesis"
/codon_start=1
/transl_table=11
/product="SimK"
/db_xref="GI:11321497"
/transl_table="MAPVAIVAGVGTGTVVAVRLAEGDVAVLCADGAGNADTVH
RVAVGERALAVETVTDVAVSQALDRVCAELGDPGLVNSGVVIACGAPGEASED
WNAVGRSLGLFTVTSRAVTDLMIRQGRGRIVTIANLIGPAEGGPHDHTVLGMAGF
TRTVALLRAFDTIVNAVAPPPVGVGQAPAAFTAADPEGSCATQAAEAVLFLVGG
ASSVTQGTQVGVGDEAAA"
847..2406
/gene="simL"
847..2406
/gene="simL"
/notes="amide synthetase; similar to novobiocin acid
synthetase NovL; involved in aminocoumarin biosynthesis"
/codon_start=1
/transl_table=11
/product="simL"
/protein_id="AAG34183.1"
/db_xref="GI:11321498"
/transl_table="MEGNEHYVQILNLTLRADPSGVALVHRDTPVIAGDLADTSAA
EAMRGSGVGVSGVIGIITDPTATLVYAANLGGATVVLFGVNAANPSDLLSABE
QGGIVAEALPAMVVVDAANLERARAREVSPVPLSGELGHDVLDLTDSPAGAFR
PDAARDGTAVTTFSSGTCGRPKGTAWSPVKADMAASARRAKATAPATVTHGN
GFVADVDVNSGGTVLLPGDFDETVLSVARYQVNLAVSAPOLYALADHPETRTDL
SSVDLFTYGTVAASPERVAEKVFGSLVCQVGTSTNIIISLLAGEHTDAGIRATV
GRLEMLVTRIDPQDERVLTGETGEVWNSPMRDHYNDDPQTARTVRDGMIRTG
DVGHLDAGYLHLHGLAGVITKNGIKVYPVAVERSLLDHPDVAEAAVFGVNSDRVE
RIHAVVLRGAGAGPEDLRQHVSSHLSPNHAPADIELRSSLPLIGFGKPKDLRLAD
AMARREATHGE"
2435..5428
/gene="simH"
2435..5428
/gene="simH"
/notes="peptide synthetase-like protein; similar to NovH of
novobiocin biosynthetic gene cluster; involved in
aminocoumarin biosynthesis"
/codon_start=1
/transl_table=11
/product="simH"
/protein_id="AAG34184.1"
/db_xref="GI:11321499"
EVDRSQPDQDSRRTGLDRLATGGEAVVVGAPPRLTLVRLRADAFQURLADRAVLDA

```



DEFINITION Sequence 42 from Patent WO03060127.  
ACCESSION AX803758  
VERSION AX803758.1 GI:38520925  
KEYWORDS Streptomyces refuineus  
SOURCE Streptomyces refuineus  
ORGANISM Streptomyces refuineus  
Bacteria; Actinobacteria; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 Farnet,C.M., Staffa,A. and Zazopoulos,E.  
AUTHORS Genes and proteins involved in the biosynthesis of lipopeptides  
TITLE Patent: WO 03060127-A 42 24-JUL-2003;  
JOURNAL Ecopia Biosciences Inc. (CA)  
FEATURES  
source  
1..18876  
/organism="Streptomyces refuineus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:23296"  
ORIGIN  
Query Match 4.4%; Score 22; DB 6; Length 18876;  
Best Local Similarity 100.0%; Pred.No. 9.2; 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0; Mismatches 0;  
QY 20 CGACCGGGCGGCCAAGGCGGT 41  
|||||  
Db 13034 CGACCGGGCGGCCAAGGCGGT 13055  
RESULT 15  
AF324838 34869 bp DNA linear BCT 14-NOV-2003  
LOCUS Streptomyces antibioticus simocyclinone biosynthetic gene cluster,  
DEFINITION partial sequence.  
ACCESSION AF324838  
VERSION AF324838.1 GI:12744817  
KEYWORDS Streptomyces antibioticus  
SOURCE Streptomyces antibioticus  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 34869)  
AUTHORS Trefzer,A., Pelzer,S., Schimana,J., Stockert,S., Bihlmaier,C.,  
Piedler,H.P., Welzel,K., Vente,A. and Bechtold,A.  
TITLE Biosynthetic gene cluster of simocyclinone, a natural multihybrid  
antibiotic  
JOURNAL Antimicrob. Agents Chemother. 46 (5), 1174-1182 (2002)  
MEDLINE 21955989  
PUBMED 11959542  
REFERENCE 2 (bases 1 to 34869)  
AUTHORS Trefzer,A. and Bechtold,A.  
TITLE The simocyclinone biosynthetic gene cluster isolated from  
Streptomyces antibioticus Tue6040  
Unpublished  
JOURNAL 3 (bases 1 to 34869)  
REFERENCE Trefzer,A. and Bechtold,A.  
AUTHORS Direct Submission  
TITLE Submitted (29-NOV-2000) Pharmazeutische Biologie,  
JOURNAL Christian-Albrechts-Universitaet Kiel, Gutenbergstr. 78, Kiel  
24118, Germany  
FEATURES  
source  
1..34869  
Location/Qualifiers  
/organism="Streptomyces antibioticus"  
/mol\_type="genomic DNA"  
/strain="Tue6040"  
/db\_xref="taxon:1890"  
/clone="cosmid 5J10"  
1337..2812  
/gene="simA7"  
1337..2812  
/gene="simA7"  
/note="involved in polyketide formation"  
/codon\_start=1  
gene  
CDS

/transl\_table=11  
/product="putative oxygenase SimA7"  
/protein\_id="AAK06782.1"  
/db\_xref="GI:12744818"  
/translation="MDAQIVVVGAGPAGLMLAGELRLAGVDVVVLERLAQPTGESRGL  
GPTARTLEVFDQRLPRLGDIETVLDGFGGLPLDIFILGPAHSGGVKGVPAVTEA  
MLDEWALELGDILRGHGLIETONGDFVEVEVATSDGQRLRTAPLVCDCGRTVTR  
KAAGDFPGTSATMEMFLGDRGCDITPRMIGETVPGGMWAAPIGNGVDRIIVCEHG  
TPARRRDEPFSSWEEAAWQRLTGODISAGEFPWLSAFGNPARQVTERRGRVLLAGD  
AAHTLHPAGGQCMNTSIQDVNVNLGWKLAATVKGTGAPGGLDLSHYTERHPVGRRLMNT  
QAQLFLSGAEMOPLRDVMSGLITTFENVRHLAQMSVGLSIRYDLGPDHPLGLGRI  
PHLELVGAGTKTSTTELLHPARGVLLDLADDAGLRETAAPWADRDPWVTVTVHGAADA  
PQYEPADALLIRPDGHVAAAGPVALSEALARSFGAPSR"  
2862..3188  
/gene="simA4"  
2862..3188  
/gene="simA4"  
/note="involved in polyketide formation"  
/codon\_start=1  
/transl\_table=11  
/product="putative cyclase SimA4"  
/protein\_id="AAK06783.1"  
/db\_xref="GI:12744819"  
/translation="MHSTLIIVARMSPRDDQDVARLFSEPDRTMPLMGTTRRQLFSY  
RGLYFHLQDFDDEHGEAEIAAKSHPSFVRISEDLPFIQAYDPDTWRSPADAMARRF  
YOWSAQ"  
3185..4453  
/gene="simA1"  
3185..4453  
/gene="simA1"  
/note="involved in polyketide formation"  
/codon\_start=1  
/transl\_table=11  
/product="putative ketosynthase SimA1"  
/protein\_id="AAK06784.1"  
/db\_xref="GI:12744820"  
/translation="MRRRVITGVGVMAPGGILGVKNFWSILLSEGRATRGITFFDPAP  
FRSRVAEIDFFPEKHGLGPKIRMDRAAQLAVIATREAVADSGLLDLDATDPRSGV  
TVGSVAAGATTGLDQRYRVVSDGRGLVDHRYAPPHLYGHVFPSSFAAEVAMASGSEG  
PTVSVTGTCTGLDGLGAVELIREGSADVMTAGATDAPITMACFADAKATPRN  
DIPAHASRPDSRNGFVLGEGHVLVLELGAARQGGARIYAEIGYASRCNAFMFT  
GLRPGREMAEIRVALDEARLNPEIDIVINAGSGTKONDRHETAAFKRSLGDHAYR  
TPVSSIKSMVSGSLGAISIEIAACVLMAESVPTTANLHDPEDCDLDYVITARE  
ARLDRLVSGSGGFGFSQSAWLTRPERNTA"  
4450..5664  
/gene="simA2"  
4450..5664  
/gene="simA2"  
/note="involved in polyketide formation"  
/codon\_start=1  
/transl\_table=11  
/product="putative ketosynthase SimA2"  
/protein\_id="AAK06785.1"  
/db\_xref="GI:12744821"  
/translation="MTTSVVTVGVGVAAPNGLCADYWAATREGRSGIGRVTFRDPSC  
YPSRLAGEVPFVAEDHRLPRLPQTDHMTLRLLVSDWALQDAGIRPEELPEVAGV  
VTASSAGGFEFGQNELRALMSKQSVYSAFVYVNTQGISIRHGLRGSGVVV  
SDQAGGLDALAQRRQIRRGSGVIVSGGIDASICFAMVGNQALTSGRNSTDDEPATAT  
PFGAAAGHVPBGEGALLVLEEGAAAGARVYRIAGYGTTFDPKPGSGREPLERR  
AIEALTDAGLAPADIDVVFADAAVFDLRIEARLVEVFGPRGLPVTAPKSGMTGRL  
NSGAGSLDVATALLARDGVIPTTINVTQDDYELDLVTAQARSARLSALVWARGHG  
GFNSALVVDQAA"  
5713..5979  
/gene="simA3"  
5713..5979  
/gene="simA3"  
/note="involved in polyketide formation"  
/codon\_start=1  
/transl\_table=11  
/product="putative acyl carrier protein SimA3"  
/protein\_id="AAK06786.1"  
/db\_xref="GI:12744822"  
/translation="MQTTEFLDLRLRDGAGSEVDVLEGLIIDVAFEDLGYESI









AUTHORS Heath,P.  
TITLE Direct Submission  
JOURNAL Submitted (05-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
On Nov 7, 2003 this sequence version replaced gi:38091229.  
COMMENT ----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BK2511C7  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 143823 bases at least Q40  
Consensus quality: 144044 bases at least Q30  
Consensus quality: 144181 bases at least Q20  
Insert size: 144260; sum-of-Contigs  
Insert size: 165629; 9.1% error; agarose-fp  
Quality coverage: 18.13x in Q20 bases; sum-of-Contigs Quality  
Coverage: 16.87x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 48654: contig of 48654 bp in length  
\* 48555 48754: gap of 100 bp  
\* 48755 139049: contig of 90295 bp in length  
\* 139050 139149: gap of 100 bp  
\* 139150 142135: contig of 2986 bp in length  
\* 142136 142235: gap of 100 bp  
\* 142236 144560: contig of 2325 bp in length.  
FEATURES  
source Location/Qualifiers  
1. .144560  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="CTD-2511C7"  
/clone\_lib="CIT-HSP-D2"  
misc\_feature 1. .48654  
/note="assembly fragment:01776  
fragment\_chain:1"  
misc\_feature 48755. .139049  
/note="assembly fragment:04387  
fragment\_chain:1"  
misc\_feature 139150. .142135  
/note="assembly fragment:04669  
fragment\_chain:2"  
misc\_feature 142236. .144560  
/note="assembly fragment:04697  
fragment\_chain:2"  
ORIGIN  
Query Match 4.2%; Score 21; DB 2; Length 144560;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 62 GCCTCTGCTCGTCTCTCT 82  
|||||  
Db 56213 GCCTCTGCTCGTCTCTCT 56233  
|||||  
RESULT 21  
AL646066  
LOCUS 189050 bp DNA linear BCT 11-APR-2003

DEFINITION Ralstonia solanacearum GMI1000 chromosome, complete sequence;  
segment 10/19.  
ACCESSION AL646066 AL646052  
VERSION AL646066.1 GI:17428714  
KEYWORDS  
SOURCE Ralstonia solanacearum  
ORGANISM Ralstonia solanacearum  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Burkholderiaceae; Ralstonia.  
REFERENCE 1  
AUTHORS Salanoubat M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,  
Arlat, M., Billault, A., Brottier, P., Camus, J.C., Catolico, L.,  
Chandler, M., Choienne, N., Claudel-Renard, C., Cunnac, S., Demange, N.,  
Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T.,  
Sigulier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M.,  
Weissenbach, J. and Boucher, C.A.  
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum  
JOURNAL Nature 415 (6871), 497-502 (2002)  
MEDLINE 21681879  
PUBMED 11823852  
REFERENCE 2 (bases 1 to 189050)  
AUTHORS Boucher, C.A.  
TITLE Direct Submission  
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS  
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,  
F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire  
INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.  
FEATURES  
source Location/Qualifiers  
1. .189050  
/organism="Ralstonia solanacearum"  
/mol\_type="genomic DNA"  
/strain="GMI1000"  
/db\_xref="taxon:305"  
170. .3325  
/gene="RSC1698"  
/note="synonym: RS02885"  
170. .3325  
/gene="RSC1698"  
/function="elements of external origin; phage-related  
functions and prophages"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transal\_table=11  
/product="PROBABLE BACTERIOPHAGE PROTEIN"  
/protein\_id="CAD15400.1"  
/db\_xref="GI:17428715"  
/db\_xref="SPTREMBL:O8XYR0"  
/translation="MRNIGYGGKGGSSSTTEAPDLSHVAARVLDLVSEGEIA  
GLVNGLSIYELGTPLANADGTLNFQNVAVDYRPGQDQAEIAPFPVSESEAVGVSL  
TAMPVRAINTQLSAVRVQLSRAATSGNNGYRVEYADLVADGALQOV  
LATAPGKTKSVYTRHRIELPRAKGTWIRVRIIPNANGSTIADVTGIESIAEVID  
AKLRYPNSALIGYRIDARQESSVPTSHLGRILIRVPSNYDPATRTYGLMGDTFKV  
AYSNFAPVFDIVLHRCGLGDRGVADWVKWALYQIGYCDLVDVDPGQGPRT  
CNCYLQORNDAYPLQDLASVFRGMAYWASNVAVADMPASYSYLFHPGNVDGRT  
YAGSARRAKTVALVNDPGDYAKRVEHVEDGDGIARYSIOQTEVTAFCGTSQSA  
QRYVGWILLTSRLTETVSPKGLDAVMPGAIIEVADPARAGTRGVRGADGT  
VVVDRPVVVEGDTLLINLPDGTAAQRTISRVDSQSLTVSADMSQPVQAARVSVESA  
DLKTQLFRVSVTESDGLTFDLSALEYNQSKFSAIDNGTRIEARPSVLPPSVQAPPT  
DVTLATYVVVDQGIATITMWAVQQAASAVAYTEWRDNGEWWTAGRTGQSLEVERN  
IYAGTYVARVRAINLADVASLPAISAEITLTGKTSPPPVVTGLIATAIVFGIRLDWAF

```

PTGLDVBTELMWSKADPRASATKGLDFAPFANTHTLMGLAAGATFFFWARLVKSG
NVGAYPSGNGVLGSSDRADILSYMGQIKTQLAGDLISAVDSIDPPMAGSGDEF
AGDGTVAGIVSTQSVLEARNIAAQVTTMOATVAQSAAVQVAQQAQVAEQGRLAA
MYTKTQIAANGRIYLAGIFGVNGKVIQSOVLIAADRPAVLHPNGNNVFTFPVVO
GQVPMDSATADGTITTSKIGDSIQSNFVAGTGWRLSKSLFENNGNGAGGRVD
TSALTQIYDANGTLRLVGVW"
/ gene 3662. .4318
/ note="RSC1699"
/ note="synonym: RS02888"
/ gene="RSC1699"
/ function="miscellaneous; unknown"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by FrameD"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="HYPOTHETICAL PROTEIN"
/ protein_id="CAD15401.1"
/ db_xref="GI:17428716"
/ db_xref="SPTREMBL:Q8XYQ8"
/ translation="MPAGLTVYGERGVQITEAYGNLALRQKSTVTPDGSRGSGTFS
ASRPVCIQSANFVALIGSSNGADWTKFSAPAQAPFTVYVDFEAPAPSGQGLQVF
KADGTLAFDGLAYLKVGSVTPPSPGAPYGAQWEASPLTAGNYAACLSYTRGTIYAV
PGDTTLFVADHYVTTTANGAGLKLQYTHRGAWNQDGGSLKIDTANPPQIILVDVSVL
"
/ gene 4387. .4878
/ note="RSC1700"
/ note="synonym: RS02889"
/ gene="RSC1700"
/ EC_number="3.2.1.-"
/ functions="elements of external origin; phage-related
functions and prophages"
/ note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by FrameD"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="PROBABLE BACTERIOPHAGE ENDOLYSIN PROTEIN"
/ protein_id="CAD15402.1"
/ db_xref="GI:17428717"
/ db_xref="GOA:Q8XYQ8"
/ db_xref="SPTREMBL:Q8XYQ8"
/ translation="MPYTDPSVLGGRLAFLDMVGFSEGTNGRQPTDRDRGYDVLVG
GGLFVSYADHPRLVLPRLGKSTAGRYQLIARYDYPRRLQLRLVDFGPAQDAIA
VQOIRREGALADIQAGRLAAATAKCNWASLPGAGYGQEQKFEALRAHYLLCGQGE
VKG"
/ gene 4878. .5186
/ note="RSC1701"
/ note="synonym: RS02890"
/ gene="RSC1701"
/ function="miscellaneous; unknown"
/ note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by FrameD"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="PROBABLE LIPOPROTEIN TRANSMEMBRANE"
/ protein_id="CAD15403.1"
/ db_xref="GI:17428718"
/ db_xref="SPTREMBL:Q8XYQ7"
/ translation="MTESSEMAVAKVGSAAAGSVIALRFLPCTWORMLSFVSSLGI
GCLAGGAANVERFLLVGSGSYTHMLAVASAVFGLAIYNNANMQIPEILTLDRRFLAKE
"

```

```

5189. .5464
/ gene="RSC1702"
/ note="synonym: RS02891"
5189. .5464
/ gene="RSC1702"
/ function="miscellaneous; unknown"
/ note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by FrameD"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="PROBABLE TRANSMEMBRANE PROTEIN"
/ protein_id="CAD15404.1"
/ db_xref="GI:17428719"
/ db_xref="SPTREMBL:Q8XYQ6"
/ translation="MLTSINQVANALIFVGLWAVLTHQVPTTGGLVLAIVNFAA
LGNIVAHGTCHSWPEVALNVALCVVWAFWRFRRSALGKGRG"
5464. .6009
/ gene="RSC1703"
/ note="synonym: RS02892"
5464. .6009
/ gene="RSC1703"
/ EC_number="3.4.-.-"
/ functions="elements of external origin; phage-related
functions and prophages"
/ note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by FrameD"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="PROBABLE BACTERIOPHAGE-RELATED SIGNAL PEPTIDE
PROTEIN"
/ protein_id="CAD15405.1"
/ db_xref="GI:17428720"
/ db_xref="GOA:Q8XYQ5"
/ db_xref="SPTREMBL:Q8XYQ5"
/ translation="MPGIKGTAAIAAMLGALVGAAGATWQAHRYDKDIATLQRD
QAEKDRQARADEAQLROERKRAGLERQLGELDARHYEELNNKATEARLAAALAG
TQLSIRARCTAPADGVSTTSGAGLGGATALLYGEDAADIQALVADADAIAAAL
RFFQGRERAIQTRDTSVDARR"
/ complement(6117. .7304)
/ note="predicted by Homology"
/ insertion_seq="ISRS018 or RS05932"
/ complement(6273. .7238)
/ gene="ISRS018"
/ note="synonyms: RS02893, RSC1704"
/ complement(6273. .7238)
/ gene="ISRS018"
/ function="elements of external origin; transposon-related"

Query Match 4.2%; Score 21; DB 1; Length 189050;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 CTACACGTGGCGCACGCCGG 26
| | | | | | | | | | | | | | | | | | | | | |
Db 57379 CTACACGTGGCGCACGCCGG 57399

RESULT 22
HUMFLNG6PD/c HUMFLNG6PD 219447 bp DNA linear PRI 17-MAY-1999
LOCUS Homo sapiens chromosome X region from filamin (FLN) gene to
DEFINITION Glucose-6-phosphate dehydrogenase (G6PD) gene, complete cds.
ACCESSION L44140
VERSION L44140.1 GI:1203968
KEYWORDS 1A gene; 2.19 gene; ABP-280 gene; DNase I-like protein
gene; EMD gene; FLN gene; G4.5 gene; G4.8 gene; G6PD gene; GDI
gene; Gdx gene; P3 gene; QM gene; STA gene; XAP-1 gene; XAP-2 gene;

```

XAP-4 gene; XAP-5 gene; XAP-7 gene; actin-binding protein; emerlin; emery-dreyfuss syndrome; filamin; glucose-6-phosphate dehydrogenase.

## SOURCE

ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS 1 (bases 1 to 219447)  
Chen,E.Y., Zollo,M., Mazzarella,R.A., Ciccodicola,A., Chen,C.-N., Zuo,L., Heiner,C., Burrough,F.W., Ripetto,M., Schlessinger,D. and D'Urso,M.

## TITLE

Long-range sequence analysis in Xq28: thirteen known and six candidate genes in 219.4 kb of high GC DNA between the RCF/GCP and

G6PD loci Hum. Mol. Genet. 5 (5), 659-668 (1996)

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

Submitted by:

Ellson Chen,

Advanced Center for Genetic Technology,

Applied Biosystems Division of Perlin Elmer Corp., 850 Lincoln

Center Drive, CA 94404 USA

and

David Schlessinger,

Department of Molecular Microbiology and Center for Genetics in

Medicine

Washington University School of Medicine,

St. Louis MO 63110 USA

e-mail: ellson@genseq.apldbio.com and davidsg@genetics.wustl.edu

Note: Gene predictions were accomplished with runs of Grail

versions

1.1 and 1.2, coupled with fasta and blastx comparisons to genbank

& non-redundant peptide libraries. Repeat analysis was accomplished

via

centroid

## FEATURES

## source

## Location/Qualifiers

1..219447

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="X"

/map="X"

4..223

/evidence=experimental

/rpt\_family="Alu-Sp or Alu-Sq"

226..503

/evidence=experimental

/rpt\_family="Alu-J or Alu-S"

641..931

/evidence=experimental

/rpt\_family="Alu-J or Alu-S"

1023..1313

/evidence=experimental

/rpt\_family="Alu-Sx"

complement(1446..1536)

/evidence=experimental

/rpt\_family="Alu-J"

3095..3426

/evidence=experimental

/rpt\_family="CpG Island"

complement(3663..3952)

/evidence=experimental

/rpt\_family="Alu-Sb0 or Alu-Sb1"

4912..5046

/evidence=experimental

/rpt\_family="Alu-J or Alu-S"

5047..5333

/evidence=experimental

/rpt\_family="Alu-Sb0 or Alu-Sb1"

5346..5628

/evidence=experimental

/genes="FLN"

complement(18984..19488)

repeat\_unit /rpt\_family="Alu-J or Alu-S"  
5632..5805  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sq"  
6363..17698  
/evidence=experimental  
repeat\_unit /rpt\_family="11Kb repeat1"  
6369..6633  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sx"  
complement(7121..7409)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sp"  
complement(7358..7432)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sb0 or Alu-Sb1"  
complement(7426..7701)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-J"  
8306..8592  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sx"  
complement(8863..9029)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-J or Alu-S"  
complement(9031..9317)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sq"  
complement(9318..9442)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-J or Alu-S"  
9486..9773  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-J"  
9996..10132  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sp"  
10159..10255  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sq"  
10271..10367  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-J"  
complement(10455..10739)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sx"  
complement(10740..10860)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-J or Alu-S"  
complement(10866..11156)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sx"  
11223..11501  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-J"  
11638..13564  
/evidence=experimental  
repeat\_region /rpt\_family="CpG Island"  
complement(11884..12163)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sb0 or Alu-Sb1"  
complement(12195..12482)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-J or Alu-S"  
complement(12668..12945)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-J or Alu-S"  
complement(18360..18648)  
/evidence=experimental  
repeat\_unit /rpt\_family="Alu-Sb0"  
complement(18984..45006)  
/genes="FLN"  
exon complement(18984..19488)

```
/gene="FLN"
/number=48
/evidence=experimental
complement(join(19301..19488,19814..20017,20101..20319,
20483..20659,21361..21493,22033..22148,22336..22473,
22633..22899,23005..23127,23224..23376,23453..23656,
23748..23909,24006..24179,24367..24495,24603..24743,
24833..24935,25068..25163,25277..25524,27701..27724,
27884..28073,28649..28805,28895..29018,29434..29604,
29696..29856,29934..30096,30182..30355,30440..31037,
31758..32030,32120..32237,32429..32598,32692..32782,
32868..33028,33111..33234,34472..34615,34709..34822,
34976..35038,35271..35407,35586..35709,35798..35935,
36473..36673,36757..36919,37011..37088,37181..37299,
37846..37993,38090..38187,38291..38539,41319..41691))
/gene="FLN"
/codon_start=1
/evidence=experimental
/product="filamin"
/protein_id="AAA92644.1"
/db_xref="GI:1203969"
/translation="MSSSHSRAGQSAAGAPGGVDTDRDAEMPATEKDLAEDAPWKKI
QNTTRWCNEHLKCVSKRIANLQDLSGLRLIALLEVLISQKKHKKXNQRTFRQM
QLENVSVALEFDRSIEKLVSIDSKAIVDGNLKLILGLLWILILHYSISMPWDEED
EAAKQTPKRLIGWIQNKLPQPIINFSDRWQSGRAGALVDSCAPGLCPDWDSWA
SKPVTNARAAQADMLGIPQVITPEEIVDPNDEHSVMTYLSQFPKAKLPKAPLR
KLNPKARAYGGLIPTGNMVKRAEFTVETRAGQGBVLVVEDPAGHKEAKVTA
NNDKFTFSVAVPVETGTHKTVLPAQOHIAKSPFVVDKSDQSKVTAQGCLE
PSGNTANKTYFEIFRAGAGTGEVVIQDPWGQKGTVPQLEARGDSYRCSYQPTM
EGVHTVHTFAGVPVPIPRSPYTVTGQACNPSACRAVRGLQPGKVRKETAQFKVTK
GAGSGELKTVTKGKEERVKQDGLGVGYFEYPMVPTGITVITWGQNIQSFPF
EYKVGTECGNKVRAMGLEGVGKGSADFVEAIGDDVTGTFVSEGVESQAKIED
DKDGSVDVRYPOEAGEVAHVLCNSDIRLSPFNADIPDOPDPHPRVKARGPGL
EKTGVAVNPAETVDVAKHGKAPLAVQVDNEGCPVEALVNDGNGTYSCSVPEKP
VHTHNVGSGVSIENPSFRNVGASHENKVKYGPVAKTGLKAHEFTYFVDCAE
AGGDVSIIGKCAPGVVGAEDIDFDIIRNNDFTVTKTPRGAAGSYTIMVLFDQA
TSTPSIRVFEVSHDASKVKAQGLSRTGVELGKETHFTVNAKAGKGLDVQFSL
TKGADVDDIIDHNDITVTKVTPVQGPVGNVTYGGDPIPKSPFVAFSPSLDS
KIKVSLGSEKVDVQKDEFTVKSAGGQGVKASKIVGPSAAVCPKVPGLGADNSV
VRLPREEPGEVETVDGVVPSPFPLEAVAPTKSVKAFGPGQLQGSAGSPARF
TIDKAGTGGGLGVGECCEAQLCDNDGTCVSYPTEPGDINILFADTHIP
GSPFKAHVFCDSKVGSGPGLERATAGEVQFQDCSSAGSALTEICEAAGLP
AEVYIDHGDGHTTIYIPLCPGATVTIKYGGQVFNPSKLOVFPVAVDTSGVCYG
PTEGIGQVFREATFTSDARALTGTGGPHVKARVANPSGNTLETYVQDRGDGMKYVE
YTPYREGLHSVDVYDGSVPSPFPVTEGDSRVRVHGPGIOGQITNKPNKFTV
ETRGATGGGLGAVEGSPSEAKMSCMNDKSGSCSVETPYEAGTYSLVNVTYGGHVPGS
```

Query Match 4.2%; Score 21; DB 9; Length 219447;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GCCTCTGTCGCTGCTTCT 82  
Dd 99830 GCTCTGTCGCTGCTTCT 99810

RESULT 23  
BX640433  
LOCUS  
DEFINITION  
BX640433 349305 bp DNA linear BCT 14-AUG-2003  
Bordetella parapertussis strain 12822, complete genome; segment 11/14.  
ACCESSION  
BX640433.1 GI:33574176  
VERSION  
BX640433.1 GI:33574176  
KEYWORDS  
complete genome.  
SOURCE  
Bordetella parapertussis  
Bordetella parapertussis  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Alcaligenaceae; Bordetella.  
REFERENCE  
1  
AUTHORS  
Parkhill, J., Sebaihia, M., Preston, A., Murphy, L. D., Thomson, N., Harris, D. E., Holden, M. T. G., Churcher, C. R., Bentley, S. D., Mungall, K. L., Cerdano-Tarraga, A. M., Temple, L., James, K., Harris, B., Quail, M. A., Achtman, M., Atkin, R., Baker, S., Basham, D., Bason, N.,

Cherevach, I., Chillingworth, T., Collins, M., Cronin, A., Davis, P., Doggett, J., Feltwell, T., Goble, A., Hamlin, N., Hauser, H., Holroyd, S., Jagels, K., Leather, S., Moule, S., Norberczak, H., O'Neil, S., Ormond, D., Price, C., Rabinowitsch, E., Rutter, S., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L., Whitehead, S., Barrall, B. G. and Maskell, D. J.  
Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica  
Online Publication  
Nat. Genet. DOI:10.1038/Ng1227-10.1038/Ng1227  
2 (bases 1 to 349305)  
REFERENCE  
2 (bases 1 to 349305)  
AUTHORS  
Sebaihia, M.  
DIRECT SUBMISSION  
Submitted (06-AUG-2003) Submitted on behalf of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Location/Qualifiers  
FEATURES  
source  
1..349305  
/organism="Bordetella parapertussis"  
/mol\_type="genomic DNA"  
/strain="12822"  
/db\_xref="taxon:519"  
110..562  
/locus\_tag="BPP3223"  
110..562  
/locus\_tag="BPP3223"  
/note="Similar to Ralstonia solanacearum probable transmembrane protein rsc0795 or rsc05042 SWALL:Q8Y195 (EMBL:ALU46061) (150 aa) fasta scores: E(): 6.4e-34, 63.3% id in 150 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative inner membrane protein"  
/protein\_id="CAE38508.1"  
/db\_xref="GI:33574177"  
/translation="MLQDALLAWLHYLAIFVILVMTABAVLLRPMQMSARSVGRLLAY  
DLRYLASALAVLATGIRLTGAKGAFTYMANPWFHAKTGLFVLIALCSIPPTLAFLS  
WKQSLSQPGFTPADADIRRAWRWVIESHLFIPLFAVLWARGIGI"  
join(122..190,248..307,350..418,485..553)  
/locus\_tag="BPP3223"  
/note="4 probable transmembrane helices predicted for BPP3223 by TMHMM2.0 at aa 5-27, 47-66, 81-103 and 126-148"  
complement(578..1729)  
/locus\_tag="BPP3224"  
complement(578..1729)  
/locus\_tag="BPP3224"  
/note="Ortholog of Bordetella pertussis (BX470248) BPP3256"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="CAE38509.1"  
/db\_xref="GI:33574178"  
/translation="MPAHLYGRILYHLRMIPLVPLPSLFSVILLASALGALLGNSA  
ALAFSGRRARAFRRRLACGLYAVLLVLAGYFAFGVAAVYRIEYERWRHRTTL  
PEARVGGIDMPAGTIRLSEVAHSPDTPYNAEPFHEVSAHGLQALRIEYIALDHPD  
TYKLTAYPLTMRVTGPTQAIQGWRCDAASPMHFSTRDRIRAECEWLAEGNRAD  
GMEVPGARLLAHGTYTDGHDVDPWQVLEPDMAVRIGGVWLAGAIIRLDAERY  
DAFERAELACPLAFGPMHYFAGTVEVSAGRWERHFGAWIPSLAGASARYAGHFDV  
ADQAVVQGRGGEVLAVPNNRAGVLRFAIAVGGNDAAAPRRACPPR"  
complement(1589..1729)  
/locus\_tag="BPP3224"  
/note="Signal peptide predicted for BPP3224 by SignalP 2.0  
HMM (Signal peptide probability 0.997) with cleavage site probability 0.556 between residues 47 and 48;  
signal-peptide site"  
complement(join(1475..1540,1604..1669))  
/locus\_tag="BPP3224"  
/note="2 probable transmembrane helices predicted for BPP3224 by TMHMM2.0 at aa 20-42 and 63-85"  
complement(1779..3089)  
/genes="gitA"  
/locus\_tag="BPP3225"



```

source
1. 5249
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
1. 5249
/gene="Ca(2+)-sensing receptor"
515. 3772
/gene="Ca(2+)-sensing receptor"
/note="this sequence comes from Fig. 2"
/codon_start=1
/product="Ca(2+)-sensing receptor"
/protein_id="AAB29171.1"
/db_xref="GI:453109"
/translation="WALYSCWILLAFSTWCTSAVGPORAKKGGDIIILGGLFPIHFG
VAVKODLSRSESVCEIRYNRGRWLMQAMIFAEERINSPLLNTMLTVRIPTDC
NTVSKALEATLFSVAQNKIDSLNDEFNCSEHIFSTIAVVGATSGISTAVANLGL
FYIPQVSASSRLSNKQFKSLRTIPNDEHQATAMADIIIEYFRNWNVGTAAADD
YGRPGIEFREAEERDICIQSELIQSDEEIKQVVEVIGTQNGKAGQIPGREF
LEPLIKEIVRRNITIGWLASAWASSLIAMPEYFHVVGTTIGFGLKAGQIPGREF
LOKHPKSVHNGFAKEPWEETFNCHLQKAGKPLVDITFLGHEGGARLGNSTPAF
RELCTGEENISSVETPMYDTHLRISYNYLVAVSYAHALQDIYTCIPGRGLFTNGSC
ADIKKVEAWQVILKHLRLNFTSNMGQVTFDECGDLAGNYSIINWHLSPEDSGIVPKE
VGYNYVYAKGKRLFNDKILKWSGFRVFNCSRDCLAGTRKGIIEGEPCCPEC
VECPDEYSDETDASACDDPDQFENENHTSCIAKEIBFLSWTEFGALTLFVLIG
IFLTAFLGVFIKFRNTPIVKA TNRELSYLLLFSLCCFSSLFFIGEPQDWTCLRQ
PARGISFVLCISILVKTNRVLLVFAKIPTSFHRKWGLNQLQFLVFLCTFMQIVIC
ALWNTAPSSYRNHELEBIIIFITCHEGSLMALFLIGYTCLLAALCFEFAFKSRKL
PENNFAKITTSMLFFIIVWISFIIPAYASTYKGFVSAVEVIALAASGLLACIFPN
KYIILFKFSRNTIEVRSTAFAKVAATLRNSVRQSRSSLGSGTSGTSPSS
ISKSNSEDPFQOQKQPOKQALSPHNAQPPRPSTPQPOQSQPPRCOK
VIFGSTVTFSLSFDPQKTAHNRNSTHTQSLKAKNDALTKHALLPLOCETDS
ELTSQGTGQGFVEDHLEMEDEPMSPALVVSNSRSFVSGGSGTVTENMLRS"

ORIGIN
Query Match 4.0%; Score 20; DB 4; Length 5249;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 25
AR012622 AR012622 5275 bp DNA linear PAT 04-DEC-1998
LOCUS
DEFINITION Sequence 1 from patent US 5763569.
ACCESSION AR012622
VERSION AR012622.1 GI:3970612
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Brown,E.M., Hebert,S.C. and Garrett,J.E. Jr.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 5763569-A 1 09-JUN-1998;
FEATURES
Location/Qualifiers
source
1. 5275
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 4.0%; Score 20; DB 6; Length 5275;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 26
AR012622 AR012622 5275 bp DNA linear PAT 04-DEC-1998
LOCUS
DEFINITION Sequence 1 from patent US 5763569.
ACCESSION AR012622
VERSION AR012622.1 GI:3970612
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Brown,E.M., Hebert,S.C. and Garrett,J.E. Jr.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 5763569-A 1 09-JUN-1998;
FEATURES
Location/Qualifiers
source
1. 5275
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 4.0%; Score 20; DB 6; Length 5275;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 27
AR078217 AR078217 5275 bp DNA linear PAT 31-AUG-2000
LOCUS
DEFINITION Sequence 1 from patent US 5962314.
ACCESSION AR078217
VERSION AR078217.1 GI:10004963
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Brown,E.M., Hebert,S.C. and Garrett,J.E. Jr.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 5962314-A 1 05-OCT-1999;
FEATURES
Location/Qualifiers
source
1. 5275
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 4.0%; Score 20; DB 6; Length 5275;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 28
AR177781 AR177781 5275 bp DNA linear PAT 17-DEC-2001
LOCUS
DEFINITION Sequence 1 from patent US 6313146.
ACCESSION AR177781
VERSION AR177781.1 GI:17920136
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Van Wagenen,B.C., Balandrin,M.F., DelMar,E.G. and Nemeth,E.F.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 6313146-A 1 06-NOV-2001;
FEATURES
Location/Qualifiers

```

```

AR028465 AR028465 5275 bp DNA linear PAT 29-SEP-1999
LOCUS
DEFINITION Sequence 1 from patent US 5858684.
ACCESSION AR028465
VERSION AR028465.1 GI:5940438
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Nemeth,E.F., Brown,E.M., Hebert,S.C., Garrett,J.E. Jr., Van
Wagenen,B.C., Balandrin,M.F. and Del Mar,E.G.
TITLE Method of screening calcium receptor-active molecules
JOURNAL Patent: US 5858684-A 1 12-JAN-1999;
FEATURES
Location/Qualifiers
source
1. 5275
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 4.0%; Score 20; DB 6; Length 5275;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 27
AR078217 AR078217 5275 bp DNA linear PAT 31-AUG-2000
LOCUS
DEFINITION Sequence 1 from patent US 5962314.
ACCESSION AR078217
VERSION AR078217.1 GI:10004963
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Brown,E.M., Hebert,S.C. and Garrett,J.E. Jr.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 5962314-A 1 05-OCT-1999;
FEATURES
Location/Qualifiers
source
1. 5275
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 4.0%; Score 20; DB 6; Length 5275;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 28
AR177781 AR177781 5275 bp DNA linear PAT 17-DEC-2001
LOCUS
DEFINITION Sequence 1 from patent US 6313146.
ACCESSION AR177781
VERSION AR177781.1 GI:17920136
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Van Wagenen,B.C., Balandrin,M.F., DelMar,E.G. and Nemeth,E.F.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 6313146-A 1 06-NOV-2001;
FEATURES
Location/Qualifiers

```







```
/product="luxR-family transcriptional regulator protein"
/protein_id="BAB69344.1"
/db_xref="GI:1582418"
/translation="MTVEDEQDEHDLGRAAPGDFELPDVSGFDLLQYRESLQGRSR
PRETSQALLRHLLHVPESADSVCVFVSPSAADVALLVAPLEERLKLKKINDIRSA
MAADFAVHSEVKEAEPVLLDVGAAIASVTSASASEKEIITVQGGSRPPELLA
ETMELNLSLLARGVQRTLYQHSIRKHLQITWDYGRQVVAAGAEIRTVTEIFERLICD
DTVAYIPNDGRANEALEIRHPALIRYLRAGFERAWEGAEPISPADPLPKAREDAIL
VRIQLLDVGHGTHDESISQRLGIGVTVAKRISTLSGRLSRSRGQGYLIARSQWMLDE
RSPAPAPPSPG"
CDS
1723..2700
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB69345.1"
/db_xref="GI:1582418"
/translation="MSSMWRPNPAELDKAVTGAERDAILDARSINDDARLGDSEAFY
AQQLRLIYGLHTLIPGPBAGBSVLLHTITRLLEAATIRAEAGTIIIPGLIDQPAE
PAAYLSWLKLAAREHRVYKHYPYREFIRDEATSDLRITYVQESVVDGRFDDLLAMQO
VGTSGAAQWEIAQNFWDGNGKPEEVHTLFNKI FEVFEIPADELERSLTAHLISG
NLAVLLCRYOYPEAVGFLGTEWLVPDRFVQVVAWERLGLPDIGIVYHRLHTVD
SQAAGWPHNVVIPAASEHMRRAITARGTLMLNSSARYLDERMPAVAV"
2732..3427
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB69346.1"
/db_xref="GI:1582418"
/translation="MPLNEQKFALEIPEVSQELDSVEDLVDYMGNGTRPKPFSQ
YRLPEGTEGTFEALLAHRDITFTFKNVKGQIRRVQPIKADFPVLIISHGIESLGL
DASEDQINVHQNRTRADGQVAGPLTPEGVHDHGHFVYIAVYNKNVAGAQIRLWEP
GRADFPWGTLPKPGHALLDDRLAHVDVTDESADGPGANRDIVIIAFSRWAERWYGE
BHDEAALADGAS"
3474..4397
/codon_start=1
/transl_table=11
/product="putative diaminopimelate epimerase"
/protein_id="BAB69347.1"
/db_xref="GI:1582418"
/translation="MGEHSTSTVVTVPASTPSASSPPAEFFERSAEFFTHALGNDVVV
VDPGSDFTFGPAAYVLLLCORHFGVADGVLTGPGLPAGAPGEPVRLMENSJGTCEK
SNGRLMHIAHTYGETSPWSDERFPLTAAGDVPVRILDLASGTVQMGMRPFGA
TGESLDVGGRRLAVTSVHNPNHVVPLAEISAGLARELGPLIAGHTRFPGRNTNVQFL
AIVDRRTAIEVYERAGAYTLASGSACAAAHAALGLTGPDEVREMPGGRIDVSI A
ADGSVMTCATERRISSGRSPVFRNKLETAR"
4432..5535
/codon_start=1
/transl_table=11
/product="putative aminotransferase"
/protein_id="BAB69348.1"
/db_xref="GI:1582418"
/translation="MTTATVTSTYPMRRVVDGDAAGRYDIDLGDNLPGCTLDLPL
GGLELYGNRGIPRLRALVADLYGTDDHVLITHGAAQALYLVNALLARPGDOVITF
RPGWQSDMPAGLGRDILDRPDLTVDEAAAVAGDPLRLIYVNTFCPTGQVR
SDADLAALHTLAARTGAYLLDDYALELKSAATLERTIYSSLSKSVGPGPLRG
WLYGPELVACTTERTLGNISGNCIETACDVLERRDRHYTERYLTGEGLVAVRA
WAARNDVAVLTPEPTGPPAFRLITGCEAPDLCCRRVLDGTGVLIMPGETILGGDGRFL
CFARDPHLISEGLRRLDPLVLRPOAVENMTSGAV"
5673..8723
/codon_start=1
/transl_table=11
/product="non-ribosomal peptide synthetase"
/protein_id="BAB69349.1"
/db_xref="GI:1582418"
/translation="MAYRLRGTLDVGAHRLALISLVRNTALRTTFRSARGTYPGVVA
PAGEVDLEIVDLADADLITARSSEAPFLDAAGPLYRFLSLRSLAESDHLVILVFHHI
NTDQWSALINRISISAYNAFVQGTQPALPELATEYEDHVAQREQGGDSWEQOLKF
WKCKLAGLVVEMPADRAPATPTONGETLTTFDPAELLHGARALAQEGASFLMVLIT
TALNVVYGRVGTQEDIPGLIPMLGRTEPELEDVVGFLINMLVRLTDLSDGPTFAELTE
RVADAMNDVQNDQFPDQIVSHVQAPARENPLFOIGTQLLGGSGGSDLGTLGLT
DEQLDVAPVRPDSLTLESADRLRFVAEYITDLPDGIIRIEMARHICINVLAAVE
DABQPI SKLAMLGEDERLLEAGRPDFFSHDPVHTVQLVAERTPHAVAAVCRGV
EMTYGELARRSGKLARHLRARGIRPGQVVGIVMDRDLDAVLGMVRAGGAYAVMDP
```

```
SHPAARDYMLQDTAAVPVITRADVAGGLPTSDEDRKVVLIIDAWAAIEAEDTPELEE
LADSETIAYVLTISGSGTRPKGVMIERSLAVFIEAYRHTEGETPEDRLAQOSSLIF
DMQCEYITALITGATIVLSPDEASSPEALGTLIRGQRYVYVGLSPILLSLGGPPY
PHLYIMGGAEVLPAELNVKNWIPGRKFVNLGPTAEAACTEYLCEHIEWKSPFPG
RPEYGRHYVDRHNLAVGPGGELLIGGEDGLAGVNLINPELTDEKFPNPFPEPTG
RVYETGLDPVQYDFIGRIQNVKRLRGLRTELGEIESALVAHPQIRMALVLLKA
DPQCKOLGVYTVLQPOAFTLARRHLGESMPYVPTAWVKLDBFPLTAKKIDR
HALPEPSYGADCTEGRAYVAPATTESEVRAEAFATLAKADOISADASFFELGGNSLOA
MRAVSRINKAFGVKINVMLYGGATVGGVAGAEIDELLAAKAGRD"
8737..9549
/codon_start=1
/transl_table=11
/product="non-ribosomal peptide synthetase"
/protein_id="BAB69350.1"
/db_xref="GI:1582418"
/translation="MPLREDGTHPLYCVHPVSGSYVYAGLAEILGGQQSYSGEAP
GFDDGAPLNSIEALAEHRLASLSARPRGYPILLGWSLGGVAYHMAQRIVAEAEV
PLVILIDATVPQPPPLPEKALLHAFRLDLVTAMGEPPEALDALITPDDVTPPAVV
TAVTAAGDEVLPSEDEEFLOERYDVFRAHAELVAYERHGFYACRVLPLKAADSLDA
RTMETLAPRVERHTITGDHDSINWAGEAFATLGLVROQLAQARSALAPAG"
9609..10901
/codon_start=1
/transl_table=11
/product="putative export protein"
/protein_id="BAB69351.1"
/db_xref="GI:1582419"
/translation="MEDTAAAGSAAAPRRGGGLRNDRPLVFWTGETLSLYCTQVTTL
ALPLAAVFDASAEELGVLPFLQVLPFLGLAMFGVMVDRVKRPMVIMGANAARVL
IGLVPALGAADVLLLYAVAGVQVLPFDVSWSYVPLVKDRHLVNAARVLS
TSSADVAGPGVAGVYVNAITAPTAMVNTYLLSIVSILAIQTPRPPAKPRRR
STRADLREGLRFPVGSRLIRATALVGCACNVFVSMILSLFLVAVRQGHLSAATIGLV
LGVAVGGLAGSALSGTVIRRVSVGKAYLVSGIFVAPLLIPAGGSKLMTVGFLA
SFLVYALSAVANVVISLRQTTITPDLNARMTAAMRTLMHGGGAGGGVPGVGLLAGAL
GLHEALMLAAVSAVMLVPIALSVPVSRLEHMPAPAPVP"
10979..12856
/codon_start=1
/transl_table=11
/product="non-ribosomal peptide synthetase"
/protein_id="BAB69352.1"
/db_xref="GI:1582419"
/translation="MSPDVEIAAAMWSPHYDEIAAMNDARELPFGGDTLPELPT
QVSRTPDPAVFRDTRLSYTELDSERDLARLLAGHGAAPORVVALTVPRSVELVVA
LNAVTKTGAAYLPDPDVPADRIYGLLADAAPVLLITHSDVETLPPGARPEGLPLV
VDAQSEPLERAPFARPPAQAADAAVITYSGTGPVGVVPHAGVNLHWMQHEY
RUTGDVLQKTPGSDVSVWBEFFPLITGATLVVAEPGCHKDPAYLAILRGERIT
VHFVSMLOKTFDEPAALCTGLRVVCSGEALPVELADRFTASLPGVLAHLYGTE
ASVDVYETPEPGATVPIGRVYWNTRLYVLDAAALPAAGETGELVAGVQALG
VYNRPLTAERFVADVPYGPGRMYRTDLARRRPGDGLDYVGRADHVKIRGLVBL
GEIEAALBHPAVSGAALVREDRPGDKRLVGVVPAVPGTVPABRLTAAITLDPY
MVFSTVVVLESPLTPNGKLDKRALPAPDAAAAPARVAPRTPQEALLCDLFAEVLGL
SETQVGVNDSFLELGGHLSAURLVSRIRAAFTDCLKVRQVFTPTVVAALACLASNG
R"
complement(13245..14198)
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein BldK"
/protein_id="BAB69353.1"
/db_xref="GI:1582419"
/translation="MQAVDGLDFQVAEGSLGVSGCGSTTGRLITRLBFTSGK
ISYRGDITHADKKLAPVRSEIQMTFODPYASINPRHTVGTKIAGMEINGIQPSGG
REKRVRELLIEIVGLNEHYNRFPHEPSGGORISIGVARALALEPKLIVADPEVSLDV
SIOQVNVLLIQDKOLGIAFVFIADHDLALVRHFSORVAVMYLGLKIVEIADRADLYEN
PRHPTFRALLISAVFEATVEDGEAREIRLTGDPVSPINPSGCRFRTRCWAATKCMAS
EAPLVQVGVGNAGHLTACHYPTKETVPAPRLSKDPPKAAV"
complement(14315..15385)
/codon_start=1
/transl_table=11
/product="ABC transporter intracellular ATPase BldK"
/protein_id="BAB69354.1"
/db_xref="GI:1582419"
/translation="MKSTATKPAERNPQATGPFISVQDLKVHFSTBDGTVKAVDGLS
FDLGRKTLIVGESGSKSVNTLITILGNPDSTA VEGSITLIDGBELNGASEKTELEK
LRGNKMANIFQDALTSLSPHTTGKQIGETRYKHTGSKAARARALEMTLRVGIPOF
```

DVRVDYDYPHQSGMQRAMIAALCDPELLADPTTALDVTVOAQIMDLKOLQK  
EFTGLIFTHDGLVADIDVVMYGRCEVGTKEVLRTPQHPYTWGLIGSMPS  
LEGFVDVPLTPIPGSPSLNPGSGCRFHRPCFAEKVEDGRCSRPPLLEITDGRGA  
ACHLTADORAFADLAASPAN"  
complement(15446..16339)  
/codon\_start=1  
/transl\_table=11  
/product="transport membrane protein BldKC"  
/protein\_id="BAB69355.1"  
/db\_xref="GI:15824194"  
/translation="MPEDPARLACGKCTPETLEVRHNNMTDSDVQVWHLMTGIF  
LGRDYSYCHDAPCLGYSTVNRRESVLGTMDFRPTASIALGSSVVFLLGGI GAGMI  
AAYKQSGALDKVAGSGLVASSMQLVYFVGTALYCFYQWILDPQPSYTFDTKPNASW  
ANGILLPWLVLISLITWANYTMRQESOLVQENEDYVTRAKGLSRNVEFRFAWRGA  
ANGILLPWLVLISLITWANYTMRQESOLVQENEDYVTRAKGLSRNVEFRFAWRGA  
MGIIVVDVAYVAILDIPRILA"  
complement(16552..18315)  
/codon\_start=1  
/transl\_table=11

CDS

Query Match 4.0%; Score 20; DB 1; Length 23451;  
Best local Similarity 100.0%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 ACCGGCGGCCCAAGGGCGT 41  
Db 11558 ACCGGCGGCCCAAGGGCGT 11577

RESULT 32  
AF512431 28990 bp DNA linear BCT 04-JUN-2002  
LOCUS Saccharothrix mutabilis subsp. capreolus nonribosomal peptide  
DEFINITION synthetase gene cluster, partial sequence.  
ACCESSION AF512431  
VERSION AF512431.1 GI:21321595  
KEYWORDS Saccharothrix mutabilis subsp. capreolus  
SOURCE Saccharothrix mutabilis subsp. capreolus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardiaceae; Actinosynnemataceae; Saccharothrix.  
REFERENCE 1 (bases 1 to 28990)  
AUTHORS Yin, X., O'Hare, T., Gould, S.J. and Zabriskie, T.M.  
TITLE Molecular Cloning and Sequence of a NEPS Gene Cluster from  
Streptomyces Capreolus Az50  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 28990)  
AUTHORS Yin, X., O'Hare, T., Gould, S.J. and Zabriskie, T.M.  
TITLE Direct Submission  
JOURNAL Submitted (13-MAY-2002) Pharmaceutical Sciences, College of  
Pharmacy, Pharmacy Building Room 203, Corvallis, OR 97331-3507, USA  
FEATURES  
source  
1..28990  
/organism="Saccharothrix mutabilis subsp. capreolus"  
/mol\_type="genomic DNA"  
/strain="Az50"  
/sub\_species="capreolus"  
/db\_xref="taxon:66854"  
<1..728  
/notes="ORF1; similar to penicillin-binding protein 4"  
/codon\_start=3  
/transl\_table=11  
/product="unknown protein"  
/protein\_id="AAM4721.1"  
/db\_xref="GI:21321596"  
/translation="HAWNALKRVPVVRQSLALVDAPAGALASATDLVTFGRMLAGA  
PPTLPDLDLMDRAVPVRAEPFMDGMLGALVYRDGAVVWGHGDTAGTSGSLR  
INFDDGTVALTWNGSGFALWRDLPELAAGLPLVADYDGFALRRRIIPPEECAGR  
FRNGDTYKVRPADCELELTVDGEPADLSLFDGLVFMKDSDTGTSQTGRFLRDP  
QRGSIEWIQIGGLARRQPEVM"  
878..8371  
/gene="scpsA"  
878..8371

CDS

/gene="scpsA"  
/notes="ORF2"  
/codon\_start=1  
/transl\_table=11  
/product="peptide synthetase ScpsA"  
/protein\_id="AAM4721.1"  
/db\_xref="GI:21321597"  
/translation="MVCAEAGVSYRELDARANKLAHLILAKGAPPERVVAIALPRSDV  
NVVARLAVLTKTGAALFVDFPAYPAERIDFMLEDAEPLLVSTAPGHRRAAHRRCADP  
PPLPAVIVTSGSTGRPKGVVTHRGLEPAFAEIAHFVDPGRVDFQVSPSFDASV  
LELMAGPAGALVVPGLLEQIADVIDGVTGTHALIPVPAATVPDRPLPTERC  
LVVAGACAPDLVARWAPGRTMYNAGPTSTVTSWSPLEPGGAPPIGRIPTGTE  
RVUDELPRPEGELIVIGVLARGVLRRLTAARFVADPFGPGSMVRTGDVVER  
APDGEAFVGRADHQVKIRFVPEPGEVEALLRHFPAVROAVVADPEPRLVAYVVG  
EVEGLREYLAATLFDYLVPSGAFVALDPTPLSPNGKLDALPAPVADPEDFVAPRI  
DAARVVAQAOEVLGVLGAHDDFALGDSILAVRALRGLGTFVAMPEHRTVBA  
LAARVPTPIPRVPHDRPLPAPQORLFLSDGTAEQNTGVLRTGLDVSRLSEA  
LDLARHARLRTEDVVGEPVQVADHGTIPLVVGDERELCAPFLDRRGPLTRAVL  
KPLGPDQHVLIHQHILVTDGKSVANVBLADLYAGVEPAAPVADPEFVADWADR  
HQDAAYVVERLGLLELDLPTDRPRPALRTAGAVHRRPLPRDLVRLTAVGRAHDA  
LFTTLTAQVLLSARARQDIAGVTVTSGRDHADLERAVGFFVRLIVLWSRDPDLP  
FTFLDOVRGTVVEAFEHEDVPFRLVABELGVEPDSPTPLVQAVVAQOPLVBRAD  
GGALACHDLPFRHARFDLVVEFWPRDLSLTLYEYNTDLFADTEVERLAEADRLHL  
DVVADPRLSILVELDPDDAVRGLAVOLAEVETLKRHEHVAADVVPVVGRL  
VAVVPPVPAALKGAGQILPAHAIPTATFVGLHRLDADLPAPPRSDGVVAPRTPT  
EAVLADFAEVLGAPKVGVRDNPFFALGDSILGILQVTRARAAGLILTSRDLFAHQV  
AALAPHVTRDLPDADQGPVTDAPLTPIQRFVLTGSRULPEQSLVLDLDFANTEADL  
ADVAALVEHDLALRMERFEADRVNRPVGGPDLRLGELVGRVRLTAHLVVDVVS  
WRIIAEDLVTLALRRPLGRKTSYLDWARRLAGHTRAGGFDDELAHAAVTCGVPTD  
RGANTYATREVTVRLSEEDTALLRDVPEVTVQNDVLLTALGRVLADTGTDRV  
VVDGHRGRELPAIDLSRTVGMTSTVVPVALDPREWGAALKSVKERLAVPERGI  
VYDGLRIAGGAPCAPVGFNYITRGWGGQRLDLTADPMRPHLLDVGVODGRL  
EVAVYAAVHDESVARLAEYAAALRAVTAHCAAPGAGRTPSDFPAKLTQEQVD  
RITGEDAYPLTPMQAGVWFGHGERGLGSDARVYFQIITFLVDVGDPSAFARAWQ  
RVVDRTPVLRSSVLEWDEPQLQVHEAEVFTFLDMSDLDEADARERLRFLEQDR  
ALGIDLSVLRMVAVVLAPEVTVVTHHLLDQWSVQVLSDLGVEPAERPPFR  
DYAWLAQAQDAERHREVLGTFDSTPLPADPAPAGPAASRRAVALSABSDR  
LYAFARHGLTNTVMQAAWALLARSSRFEHVEPLTRLKAWSAVADGDLDFDSAVVP  
TLPVVAVDGTVPVWMLRQAAQASRRFEHVEPLTRLKAWSAVADGDLDFDSAVVP  
ENYPVGLGDMGRIEAEVETTSPLSATVPSQRLVGLVPEVPAFAAQAARTP  
LKALLVGAEDPERPVARVPMLSAERQVLDVWNGDAEFFVTTTIPERFAAQAARTP  
DAIVSFAHEHLYVELDERSNRLAHVLAAGAPERLIVALLFETDALVAVLGLVK  
SGAAYLPDIPYAPRIAGTIACKPVVLRDLPLDLSGYPTTAPRVELHPENSAVVI  
TSGSLPKPKGVVPHSNVIRLFAATRWGVDENDVWELFHSYADSVSWELWGLLH  
GRLVVSGETSRSPREFARVREGVTVPNQTSFAYQLIPEQDESERVIFAGBPLD  
LDKTDWKGAGTMNMGITETTVTVTVADQSGVPMPLRVYLDDEDEPEVPPV  
VGMVYAGPGARGVLDPLRGTASFVAVNPFVAGPGRMYRSGDLGWMDGTILCLGR  
DHQVIRGFVRLGEIEVLAAPVAVVLAVERLVAVVYVPGHVAVADLRHSA  
AALPDYVPAVYALDRLLPLNANGKLDRAALPAPERDAVTSTYVAPRTETRAAEI  
WAEVLVDVRVGVDESFFALGDSIRSLTSTKAAFVDVLSPRDVLARTVSHLAGL  
VEELVLALEALADREA"  
8375..26956  
/gene="scpsB"  
8375..26956  
/gene="scpsB"  
/note="ORF3"  
/codon\_start=1  
/transl\_table=11  
/product="peptide synthetase ScpsB"  
/protein\_id="AAM4723.1"  
/db\_xref="GI:21321598"  
/translation="MTTSREDRISALPAHLRELRKRLAGRAAPADVIIPRAGHDGPPF  
LSAQRRMLFTHQVAFETDYNALALRGLGELVDRLTAALRELVRHRESLRTTFPEVD  
GSPAQVVRVRLDVPVVDADDLAALRAERPFDLADGFGFVALLVRHAGDRLVLLLT  
SHIIVDGSILGVLTEELGALYGEALPPTDLRYGYDVAQQDRPADLDYWRDLRAGV  
RTPDLTPDRPRAERSSAGATETVPPLASLAELARTHTTFLVLLTACQALLA  
RYTGQDDIAGVTTCGRNPELQRLVGFVGTIVLSTVDTASPADLLAGTBMGVLD  
AFADBPDPKVVVAGVRDPRDPTPLFDVMMVQNACTALPALPGLVEHEHLPTRH  
ANWLDSPFAETGDGIGVIEYSTDLDFAATVRLCHLLILILARVVAEPDLAEVPP  
LILGDSFSPGPAAEVDPDTFFALFEQAARTPDATLVCGDDVLVYAEAEANRLA  
HRLVADGVPEQVVALSPRGVDVAVVGLVKAGAVYLPDPSLAGAACGACHGIG  
RGAGRPGCHDRPASHAAGPQGTGRVYVITSGSTGTPKGVVHRNLVNLVNH  
RNDFASQPLKVALTVFSFDTSWEGVLVMDAGHELHVLITDEVALEPALTALVAVYVREHI

gene

CDS







```
/gene="B10K17.060"
/number=1
complement(16986..17046)
/gene="B10K17.060"
/number=1

Query Match
Best Local Similarity 4.0%; Score 20; DB 8; Length 77292;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 GGCAGCTCCGCCGACGCT 233
Dn 29442 GGCAGCTCCGCCGACGCT 29461

RESULT 34
AF497474
LOCUS
DEFINITION
Aegilops tauschii leucine-rich-like protein gene, partial cds; seed
globulin (Glo-2) gene, complete cds; HMW-glutenin (Glu-Dty) gene,
Glu-Dty-r2 allele, complete cds; and HMW-glutenin (HMW) gene,
HMW-Dtx2 allele, complete cds.
ACCESSION
AF497474
VERSION
AF497474.1 GI:21779916
SOURCE
Aegilops tauschii
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Aegilops.
REFERENCE
1 (bases 1 to 102842)
Anderson,O.D., Rausch,C., Moullet,O. and Lagudah,E.S.
Characterization of a wheat D-genome BAC containing two paralogous
HMW-glutenin genes: distribution of genes and retrotransposon
clusters
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 102842)
Anderson,O.D., Rausch,C., Moullet,O. and Lagudah,E.S.
Direct Submission
JOURNAL
Submitted (28-MAR-2002) ARS, U.S. Department of Agriculture, 800
Buchanan Street, Albany, CA 94710, USA
FEATURES
source
1..102842
/organism="Aegilops tauschii"
/mol_type="genomic DNA"
/db_xref="taxon:37682"
/clone="BAC BHMWglu-4"
/notes="CSIRO accession number AUS18913"
complement(<1..>1959)
/product="leucine-rich-like protein"
complement(<1..1959)
/codon_start=1
/product="leucine-rich-like protein"
/protein_id="AA077579.1"
/db_xref="GI:21779917"
/translation="MRSWLLRPLALLTTTALLNTSTSSISTADHLLPALLSFKS
LITDPLGALSSWTINSSNGSTGFCSTGKSRTHPGHVALRLOGIGLSTISP
FLGNLSRLVDLSNNKLEGGIPPSLGNCFALRNILNLSFNLSNSAIPAMGNLSKLVV
LSIRKNNISGTIPESFADLATVTFSTASNVHGGIPPLGNLTKDLNVDNMMSG
HVPFALSKLNLRLFLTLNNOGLIPVLFNMSLSLERFDFESNQLSGSLPDIDGSTL
PNLKEFLFYNKFGQIPSSLNLSISLERIVLHGNRFRHGRIPSNIGQNGCLTVFVIGK
NELQATSRDDELTSANCSLSIVDIQLANLGLIPNLSISOKLETQLVGNQI
AGHPTIGTRVYKLVLEFADNLFTGTIPSDIGKLSNRLNLFQNRVHGRIPLSLGN
MSQNLKLSNNNDEGSIPTAFGNLTSLIDLSNLSLGGQIPERWMSISLAVPLNL
SNLIDGPTITHVQGVNLAINLSSNLSAIPNTGLSCIELOFLYLQGNLLHGQIP
KEFMALRGLEEDLNNLNSGVPFLESFOLLKLNLSLNFQNSGVPDPTGIFNSASI
VSLTSGMLCGSPVFFHPACPYLAPDKL"
<3320..>4000
/gene="Glo-2"
<3320..>4000
/gene="Glo-2"
/product="seed globulin"
3320..4000

gene
mRNA
CDS
```

```

/gene="Glo-2"
/codon_start=1
/product="seed globulin"
/protein_id="AA077580.1"
/db_xref="GI:21779918"
/translation="MGRFVFVFAFLAALVAGSAOQVLEQSLTDACRCRGEVQKFFILA
CROILLEQQLTRAGAGVGPLFAHMGARECCRQLESVSRCCCAALRGAMRVDEQ
SNPFLGEGRHSGSERQKQEGSGESTPEQRQEVGGQYSGSETGGGQGGGHHGVT
VTRGQORQGVLCRKPQRQEGSGEGSAQOQPAQGRVRLTKVRLFTACRIEPRQES
VFTADQY"
<7856..>9730
/gene="Glu-Dty"
/allele="r2"
<7856..>9730
/gene="Glu-Dty"
/product="HMW-glutenin"
7856..9730
/gene="Glu-Dty"
/codon_start=1
/product="HMW-glutenin"
/protein_id="AA077581.1"
/db_xref="GI:21779919"
/translation="NAKRLVLFVAVVIALVTTARGEASRQLCERELQESSLEACP
OYVDQOLAGRLPWSGLMRCQQLRDVSAKRSVAVSOVAROYETVVPKGGSYFP
GETTELQQLQGIFFWGTSSQTQGYVPSYTSRQSVYPGQASFPQGGQYPTSPQHT
PGQSQWYIYPTSLQPGQGGQIGKQGGQYIYPTSLQPGQGGQIGKQGGQYPTSPQHT
QGRQPVQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
QPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
LOQPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
YPTSLQPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
GHPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
HCTSLQGTGGAQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
QPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
"
9937..11450
/rpt_family="TAGY-1"
<61216..>63753
/gene="HMW"
/allele="HMW-Dtx2"
<61216..>63753
/gene="HMW"
/product="HMW-glutenin"
61216..63753
/gene="HMW"
/codon_start=1
/product="HMW-glutenin"
/protein_id="AA077582.1"
/db_xref="GI:21779920"
/translation="MAKRLVLFVAVVIALVTTARGEASRQLCERELQESSLEACP
ACQVMDQLRDISPECHPVVSVAGYQQIIVVPKGGSYFGETTTPQQLQRIIF
WGIPALLKRYVTSYTSRQSVYPGQASFPQGGQGGQGGQGGQGGQGGQGGQGGQGG
PGQIQQPAQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
QGGYPTSLQPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
SQPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
SQPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
QGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
QGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
QPPAQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
QGGYPTSLQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
PQGLQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
QGGYTSYVSVHEQAASLKVAKAQAQAAQLPAMCRLEGGDALASQ"
QGGYTSYVSVHEQAASLKVAKAQAQAAQLPAMCRLEGGDALASQ"

ORIGIN

Query Match
Best Local Similarity 4.0%; Score 20; DB 8; Length 102842;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 GCCGACCACCTACCACCACC 298
Dn 54828 GCCGACCACCTACCACCACC 54847
```



```

RESULT 35
AL646071/c
LOCUS
DEFINITION
    203050 bp DNA linear BCT 02-SEP-2002
    Ralstonia solanacearum GMI1000 chromosome, complete sequence;
    segment 15/19.
ACCESSION
    AL646071 AL646052
VERSION
    AL646071.1 GI:17429627
KEYWORDS
    Ralstonia solanacearum
SOURCE
    Ralstonia solanacearum
    Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
    Burkholderiaceae; Ralstonia.
REFERENCE
    1
    Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,
    Arlat, M., Billault, A., Brottier, P., Camus, D.C., Cattolico, L.,
    Chandler, M., Choisme, N., Claudel-Renard, C., Cunnac, S., Demange, N.,
    Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T.,
    Siguier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M.,
    Weissenbach, J. and Boucher, C.A.
    Genome sequence of the plant pathogen Ralstonia solanacearum
    Nature 415 (6871), 497-502 (2002)
    21681879
    PUBMED
    11823852
    2 (bases 1 to 203050)
    Boucher, C.A.
    Direct Submission
    Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
    Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
    Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
    BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
    Dausset-CRPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
    118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
    URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, et
    Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
    F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
    INRA, BP27, F31326 Castanet-Tolosan Cedex
    Christian.Boucher@toulouse.inra.fr
    Location:Colouise.inra.fr/R.solanacearum.html.
    http://sequence.colouise.inra.fr/R.solanacearum.html.
FEATURES
    source
    1..203050
    /organism="Ralstonia solanacearum"
    /mol_type="genomic DNA"
    /strain="GMI1000"
    /db_xref="taxon:305"
    complement(180..1025)
    /gene="RSc2606"
    /note="synonym: RS00854"
    complement(180..1025)
    /gene="RSc2606"
    /functions="elements of external origin; plasmid-related
    functions non-bacterial functions"
    /note="Product confidence : putative
    Gene name confidence : hypothetical
    predicted by Codon usage
    predicted by Homology
    predicted by Framed"
    /codon_start=1
    /evidence=not experimental
    /transl_table=11
    /product="HYPOTHETICAL REPLICATION INITIATOR AND
    TRANSCRIPTION REPRESSOR PROTEIN"
    /protein_id="CAU16313.1"
    /db_xref="GI:17429628"
    /db_xref="SPTREMBL:Q8XW69"
    /translation="MSSPPQQAALPQREQLDLFRALPDMAPRDSQDLMAPFFSLAKS
    RRTAIDFRAGGITIRVEGTAERGIATWDADLIWAASQIVAEIRADAGIRPSWIRAT
    PYELIFEGSTINDYQRLKALDRLOSTSIATSIETTRGRLHFRFSWINEWRELAD
    ASKPLGIELIDFWFAGVLDALVLTIDPAVERLKGGERLWLYRLVRKHGKRGHEHG
    WQDFDRLHHRKSGSAARFSDFAVDLRALVQRSLPGVVLGIEMPDGQELLFRPVL
    HTARG"
    complement(1052..1327)
    /gene="RSc2607"

```

```

/note="synonym: RS00855"
complement(1052..1327)
/gene="RSc2607"
/function="miscellaneous; hypothetical/partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAU16314.1"
/db_xref="GI:17429629"
/db_xref="SPTREMBL:Q8XW68"
/translation="MRPAPLRPAAPAFQORYLTNDDEAAEYLRLSPTLEKORVVG
GPKFRKGRVMYAVSDLDADWADQRSYEATSDEYAEHAGDYRDGR"
complement(1442..2212)
/gene="RSc2608"
/notes="synonym: RS00857"
complement(1442..2212)
/gene="RSc2608"
/functions="miscellaneous; hypothetical/partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHETICAL/UNKNOWN PROTEIN"
/protein_id="CAU16315.1"
/db_xref="GI:17429630"
/db_xref="SPTREMBL:Q8XW67"
/translation="WTRDSAEHWYPTAAIYLVHLGDGPALAWEYLRNPDYRDLWR
RRPDAQWGLLELPALDARDAHPAFPDHDAVVOLYPDADPPPAHAFEFWRVP
GRKQIHDKRLVLSVHPGCGCIRLALAGLEDGMAIYATACATPCARYRTLTAE
DALATATVAAPTAAARSPTAAVLELTLQALDALTLAGASIREVAVGLFSADAVAD
WHDQSAFARVRLVRREGEALMRGGYRLAQLELLPLQ"
2213..8108
/note="ACUR or RS00937
predicted by Framed"
/evidence=not experimental
complement(2557..3150)
/gene="RSc2609"
/notes="synonym: RS00858"
complement(2557..3150)
/gene="RSc2609"
/function="miscellaneous; hypothetical/partial homology"
/notes="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE LIPOPROTEIN"
/protein_id="CAU16316.1"
/db_xref="GI:17429631"
/db_xref="SPTREMBL:Q8XW66"
/translation="WTERPLSCFYGNPHLRIPTEGNRIKRVFPASDNSASVLPFFCGC
DGLHINASTRMHRYGSTVLCFQTRAHPRIYGVFVRRLPMTLPVTADARLLAHGAAL
AAGQRDLPLFVVRVLTFTDAHVTWLLVSLDPAQDGTAYGLIDLGISLGTVKLSLIA
GVGPRDLFVRDRYFQAHLHLLSEYVRLAQENAITD"
3223..3513
/gene="RSc2610"
/notes="synonym: RS00859"
3223..3513
/gene="RSc2610"
/function="miscellaneous; not classified regulator"
/notes="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"

```

```

/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD16317.1"
/db_xref="GI:17429632"
/db_xref="GOA:Q8XW65"
/db_xref="SPTREMBL:Q8XW65"
/translation="NAKYSIAKALKTVRKARGLSQEAFFSDVSSRTYMTSLERDLKSP
TIHKUTEJCEVDMVDFLLTLTAFAGDSTRKADELLAQVRQLEAVLNSDGD"
3522..4085
/gene="RSC2611"
/note="synonym: RS00860"
3522..4085
/gene="RSC2611"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHEtical/UNKNOWN PROTEIN"
/protein_id="CAD16318.1"
/db_xref="GI:17429633"
/db_xref="SPTREMBL:Q8XW64"
/translation="MAYRNGNTAFVVAEPFHPSSLGANATKDFQYNTLRMWKGADA
TFPVDSDHKTVSRGSDWEITLKRLEBRSSKNIIILSSITNSRALREEDYD
GINDQGLPVIVYPTAKESLLVDGSLQAVKNLGNLVPFRDSMSKIPTLHPVNNK
EIKKSLDQGEFVMSGKAPDIYRNP"
4107..4955
/gene="RSC2612"
/note="synonym: RS00912"
4107..4955
/gene="RSC2612"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHEtical TRANSMEMBRANE PROTEIN"
/protein_id="CAD16319.1"
/db_xref="GI:17429634"
/db_xref="SPTREMBL:Q8XW63"
/translation="MIHMAKVNFDRVIRKKEITAVVSGTSLSFIVIFVDVPAEWK
LKAGLAFALITLVYVIMLSNNLSIDINVGSDVAIKTGDIFQPGKLAIFNEY
FDQVDNKKIIGETSLNGVFIQKHLGVPISELDRIEYAFDESEVLEENARKYKKK
RYOIGTICLYKVDLTAFSPFDENKALLTMEYLEFLINFDKVNNVVAQSVSTTI
FGSGITRIKGHKNSDEDLKMLTFRISEMRKYPAKLTVIHKKIDQINLLDIK
SARNGV"
complement (5305..5619)
/gene="RSC2613"
/note="synonym: RS00913"
complement (5305..5619)

gene
CDS

Query Match 4.0%; Score 20; DB 1; Length 203050;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 CCTCTCGCGCGGTGCGG 282
|||||
Db 20919 CCTTCTGCGCGGTGCGG 20900

RESULT 36
AE016925/c
LOCUS 208524 bp DNA linear BCT 14-OCT-2003
DEFINITION Chromobacterium violaceum ATCC 12472 section 16 of 16 of the
complete genome.
ACCESSION AE016925 AE016825

```

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AE016925.1 GI:34105514  
Chromobacterium violaceum ATCC 12472  
Chromobacterium violaceum ATCC 12472  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Chromobacterium.  
1 (bases 1 to 208524)

REFERENCE  
AUTHORS  
CONSTRM  
TITLE

Brazilian National Genome Project Consortium  
The complete genome sequence of Chromobacterium violaceum reveals  
remarkable and exploitable bacterial adaptability  
Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)  
22822880  
14500782

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

2 (bases 1 to 208524)  
Vasconcelos, A.T.R., de Almeida, D.F., Almeida, F.C., de  
Almeida, L.G.P., de Almeida, R., Goncalves, J.A.A., Andrade, E.M.,  
Antonio, R.V., Araripe, J., de Araujo, M.F.F., Filho, S.A., Azevedo, V.,  
Batista, A.J., Bataus, L.A.M., Batista, J.S., Belo, A., vander Berg, C.,  
Blamey, J., Bogo, M., Bonato, S., Bordignon, J., Brito, C.A.,  
Brocchi, M., Burity, H.A., Camargo, A., Cardoso, D.D.P.,  
Carneiro, N.P., Carraro, D.M., Carvalho, C.M.B., Duran, N., Fagundes, N.,  
Falcão, C.L., Fantinatti, P., Farias, I.P., Felipe, M.S.S.,  
Ferrari, L.P., Ferro, J.A., Ferro, M.I.T., Franco, G.R.,  
Freatas, N.S.A., Furlan, L.R., Gazzinelli, R.T., Gomes, E.A.,  
Goncalves, P.R., Grangeiro, T.B., Grattapaglia, D., Grisard, E.C.,  
Guimaraes, C.T., Hanna, E.S., Hungria, M., Jardim, S.N., Laurino, J.,  
Leoi, L.C.T., Passarella, L., Lima, A., Loureiro, M.F., Lyra, M.C.P.,  
Macedo, M., Madeira, H.M.F., Manfio, G.P., Maranhao, A.O.,  
Martins, W.S., di Mauro, S.M.Z., de Medeiros, S.R.B., Meisner, R.V.,  
Menck, C.F.M., Moreira, M.A.M., Nascimento, F.F., Nicolas, M.F.,  
Oliveira, J.G., Oliveira, S.C., Paixao, R.F.C., Parente, J.A.,  
Pedrosa, F.O., Pena, S.J.D., Ferreira, J.O., Ferreira, M.,  
Pinto, L.S.R.C., Pinto, L.S., Porto, J.I.R., Potrich, D.P.,  
Neto, C.E.B., Reis, A.M.M., Rigo, L.U., Rondinelli, E., dos  
Santos, E.B.P., Santos, F.R., Schneider, M.P.C., Senanez, H.N.,  
Silva, A.M.R., da Silva, A.D.C., Silva, D.W., Silva, R., Simoes, I.C.,  
Simon, D., Soares, C.M.A., Soares, R.B.A., Souza, E., Souza, K.R.L.,  
Souza, R.C., Steffens, M.B.R., Steindel, M., Teixeira, S.R.,  
Urmeyvi, T., Vettore, A., Wassen, R., Zaha, A. and Simpson, A.J.G.

TITLE  
JOURNAL

Direct Submission  
Submitted (22-JAN-2003) Labinfo, LNCC - Laboratorio Nacional de  
Computacao Cientifica, Rua Getulio Vargas 333, Petropolis, RJ  
25651070, Brazil

FEATURES  
source

Location/Qualifiers  
1..208524  
/organism="Chromobacterium violaceum ATCC 12472"  
/mol\_type="genomic DNA"  
/strain="ATCC 12472"  
/db\_xref="taxon:243365"  
complement (96..563)  
/gene="pile2"  
/locus\_tag="CV4209"  
complement (96..563)  
/gene="pile2"  
/locus\_tag="CV4209"  
/note="identified by sequence similarity; putative; ORF  
located using GeneMark/Blastx/COG2165"  
/codon\_start=1  
/transl\_table=11  
/product="type IV pilin"  
/protein\_id="AAQ61869.1"  
/db\_xref="GI:34105515"  
/translation="MKKQKQOQFTLIELMIWVIGILAAIAPAYODYTKRVRVSE  
GLADAARAKTAVTEYATNNWASGTPENTTYGLAASITGNSVSSINLVANGVQIA  
INSTVNSGALVDLIVTVSAGSVQWVCTTSGTTALASASQLSANWPTCRQ"  
838..2184  
/gene="bioA"  
/locus\_tag="CV4210"  
838..2184  
/gene="bioA"  
/locus\_tag="CV4210"

gene

CDS

gene

CDS





```
/codon_start=1
/transl_table=11
/product="putative two-component system response
regulator"
/protein_id="BAC72393.1"
/db_xref="GI:29608339"
/translation="MAIRVLMVDQVLLRTGFRMLAQPDMFVBAAGDVEALQVL
RSTVDVLMVDMRPKLDGVETTRICSEPNPKVLIITPDLDEYAFSGLKAGSGF
MKDVPPEGLLTARSVHSGDAVAPSTRELLDRFAPMLPSAGKEPQKLEBLRTDR
EREVMVLVAQGLSNGEIAARLVISEATVKTHVGRLLIKLGLRDRVQVVVLAYETGLVR
AGGQT"
8125..8757
/note="SAV4682"
/codon_start=1
/transl_table=11
/product="putative TmrB-like protein"
/protein_id="BAC72394.1"
/db_xref="GI:29608340"
/translation="MLLWINGPFGGKGTQTAHEIQRRLPGSVICDPEHVGFLRRMLP
PGLANFODLASWQGVVEVDLALRDHGVVIAPMTITDSCFAQTGRLRELGHDV
RHEALLARETVLRLRERGRHLLRYVAGDAPLRGSMWAGQLDHCLERLRBPEFG
EHLWTDQSTVTKTADRIAAGLITLAPNTEGALRRLRQVRVGVKHMREF"
complement(8764..9888)
/note="SAV4683"
/codon_start=1
/transl_table=11
/product="putative AtrE-family transcriptional regulator"
/protein_id="BAC72395.1"
/db_xref="GI:29608341"
/translation="MSVCIDITGLPAERIVPETSPLAEHALALAEFGHPHATQGW
ATATSAALPDLALSEADFLWSTFSDIFMFAGVAGGVRPGATLAEDDILDRLL
DDEEFTVAALPESCGSQAAGLPSLFDRAHIALELAASRQGLQFTQRLADPP
TVRAVRLLEDCEAFADTWRRVHPOLAADARKHTELLRHKGLEALTSVAALS
NTAGRTLEALDGAHGAATDPAVGGVTEVPVSGHMPHLLVHAGSWRPVHVVP
AAPEPSAASVELLKLRLALAHVRLRLCRIAAPYTTGELAHADLSAPEVSRHL
TVLKAGLITTRRGYVHLQDVTVAKLGSDLEGLIR"
complement(10190..10318)
/note="SAV4684"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC72396.1"
/db_xref="GI:29608342"
Query Match 4.0%; Score 20; DB 1; Length 299800;
Best Local Similarity 100.0%; Pred.No. 65;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 GCGGCTGATCGAGCGCGAGA 256
|||||
Db 136409 GCGGCTGATCGAGCGCGAGA 136428

RESULT 38
AP005939
LOCUS
DEFINITION AP005939 300900 bp DNA linear BCT 12-SEP-2003
Bradyrhizobium japonicum USDA 110 DNA, complete genome, section
5/31
ACCESSION AP005939 BA000040
VERSION AP005939.1 GI:27349336
KEYWORDS
SOURCE
ORGANISM
Bradyrhizobium japonicum USDA 110
Bradyrhizobium japonicum USDA 110
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
1
REFERENCE
AUTHORS
Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,
Sasamoto,S., Watanabe,A., Idesawa,K., Iriiguchi,M., Kawashima,K.,
Kohara,M., Matsumoto,M., Shimo,S., Tsuruoka,H., Wada,T., Yamada,M.
and Tabata,S.
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110
DNA Res. 9 (6), 189-197 (2002)

TITLE
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110
DNA Res. 9 (6), 189-197 (2002)

JOURNAL
```

```
22484998 MEDLINE
12597275 PUBMED
2 REFERENCE
AUTHORS
Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,
Sasamoto,S., Watanabe,A., Idesawa,K., Iriiguchi,M., Kawashima,K.,
Kohara,M., Matsumoto,M., Shimo,S., Tsuruoka,H., Wada,T., Yamada,M.
and Tabata,S.
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)
DNA Res. 9 (6), 225-256 (2002)

TITLE
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)
DNA Res. 9 (6), 225-256 (2002)

JOURNAL
MEDLINE
22485002 PUBMED
12597279 PUBMED
3 (bases 1 to 300900)
AUTHORS
Kaneko,T.
Direct Submission
Institute (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research; 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp/rhizobase/,
URL:http://www.kazusa.or.jp/rhizobase/,
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
Location/Qualifiers
1. 300900
/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
103..756
/gene="blr1086"
103..756
/gene="blr1086"
/note="ORF ID:blr1086
Tetr family"
/codon_start=1
/transl_table=11
/product="transcriptional regulatory protein"
/protein_id="BAC46351.1"
/db_xref="GI:27349337"
/translation="MKSPSDETTLTNTYVRIILMPKRSKDALDAGLKVMFRTGY
HGTSVRDVTAAAGAPQSGFTNHRSPASEVLDORYFDVTGLVAEALDITLTPRA
RLKRYLDITIGLEADGYRGCLIGDLSLEATGSGSEMRLTRITEIFABWRVFAACIA
EAQARGEIASDPEPGLADFLIASWQGAIIIRMKVDRNPKALERFKTAIPQTVEKELT"
753..1685
/gene="blr1087"
753..1685
/gene="blr1087"
/note="ORF ID:blr1087"
/codon_start=1
/transl_table=11
/product="haloalkane dehalogenase"
/protein_id="BAC46352.1"
/db_xref="GI:27349338"
/translation="MSKPTIEIRRAPVLGSSMAYRETQAQDAPVVLFLHGNPTSSHI
WRNILLVGPVAHCTAPDLIGFGQSGKPDIAVRFDFHRYLDAFIQRGVTSAYLVQAQ
DMGTALAFILAAARRDPFVRGLAFMEFTPMPTWODEFHTVEAEQDHAFAAFAVFRKF
RPFGEHMLFANAFVERVLPGGIVKLGDEMAFYRTEFTPESSRPVLAAPPRLP
IAGEPADVDAIGSAHAALASYPKLLFTGPGALVSPFEAFERFAASITRCALIRLG
AGHYIQEDHAALIGRSVAGWIAIEAVRPQAA"
1941..2372
/gene="blr1088"
1941..2372
/gene="blr1088"
/note="ORF ID:blr1088
putative GNAT family acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="BAC46353.1"
/db_xref="GI:27349339"
/translation="MPVWNEDPAALDMNLSALYTAIPGNKTPADLALVFGNSMFC
AFAYDQRLVAGRALGDCAYLCDAVVRPDCQGLGRE:IERLLAECRGRKII
LYAVPGKEGFEVRLGFRMTTAMAFIDDDQASAHQRGYLTAP"
2795..3850
/gene="blr1089"
2795..3850
/gene="blr1089"
```

CDS	2795..3850 /gene="blrl1089" /note="ORF_ID:blrl1089" hypothetical protein /codon_start=1 /transl_table=11 /protein_id="BAC46354.1" /db_xref="GI:27349340" /translation="MLEARRAMTFRLKQVHLKGVLSVAIGIACVVLVHMLRGI DFNEVEAIKSTEPSQIAWALFVAAGYFTLFDYFAVRAIGHVPRINALAFT SYSHGNVASYFTGAVRIYSYAGLNAIDVAKICTFLAGLTFWIGNAAYLGLGISY HPEAASIDQLPWLNLALMIIVGLGVVWVWTPRVGVGWPTVVLPGSLTLL QJAGIDIDGFCALANVLPDEPNLGFVVAVIVFSATLLGFASHPGGLGVDAAM LVGLWMDRELLGCMLLFRVLYLSPFVLSVILLTFREVIIGARSKRLQQAALKLDLP GPAREAAVYRERSDGT" 3910..5208 /gene="phor" 3910..5208 /gene="phor" /note="ORF_ID:blrl1090" /codon_start=1 /transl_table=11 /product="phosphate regulon, two-component sensor histidine kinase" /protein_id="BAC46355.1" /db_xref="GI:27349341" /translation="MAIDAFSSPTAQPSDRLHSTILLIAAALSVVVSIGLSVL HAVAFLCTAAALIPWRLHDTAASDDVRRINPVSAAVAAGMPDPVLLDRAG RVTHLAAQAALPALRNLQAFLRSPBIIITALRESIATTPRRATYLDHVPVDRW MELIITPVPTTGGADKCMWLTFFDQTPLRREMRADFVANASHELTPLAALSG FIDTLOQAKDDPKARERFGIMHNOATRMARLIDLLSLRVELSAHVRPDTLVDLL PIRVADGLEPLARERQVEVHLPEPPVMIAGDREELARLPENLLENALYKAGSGG RVIVLSTGAATDGTQIRVWVRDFGCIAPDEHPLPLTERFYRYVDVGDLSRLQGTGILG LSLVHLINRHRGRLLESVPKQATFACFPQIRPPVTV" 5341..6366 /gene="pstS" 5341..6366 /gene="pstS" /note="ORF_ID:blrl1091" /codon_start=1 /transl_table=11 /product="ABC transporter phosphate-binding protein" /protein_id="BAC46356.1" /db_xref="GI:27349342" /translation="METSMLPXTIIVAGLVAATTAADINAGATPPPIYSKW ADAYKETGNGINTQSIGGGIKIOIAKTVTFGSDAPLKAQLEKDKGLVQWPMVG AIVPVNIEGVKPGEMVFDGETLASIYLGKITKWDAAIKGNLNLKPLSDAITVRR SDGSGTTFNFNYLSKASADWKSQVGETAVWPVGAKNEGVSNGIKTKNSIGY VEYAYAKNLTYYTLVYNKAGKPVQPTVEAFQAAASNADWAKAPYVYVILTDQGEKS WPTAATFILMKDQATKASQEAALKFRWAFKNGKAAEELDYIPMGVVTLEKT WAAEIKS" 6487..7485 /gene="pstC" 6487..7485 /gene="pstC" /note="ORF_ID:blrl1092" /codon_start=1 /transl_table=11 /product="phosphate ABC transporter permease protein" /protein_id="BAC46357.1" /db_xref="GI:27349343" /translation="MAEMAVQSDVIDDAGPYDRAKALSAFKLGDVTFYVITLSAISV LLILGGLIVSLIVGAPFAIKYGFSLFWTQWAPSDPPVLGALGPMYGTLYTSIAM LIAIPVGLIGALFELCPQWLRIPIGMALIELLAGIPSIYIGMWGFFVLGPLEANTFQ PFMIKIFDGVPLVGLAFAGPSYLSLNAALIAIMVLPFTISVDFKTPVPLVKE RAYGVCTTWEVRSWVTPYTRGVIGVGMALGALGETMAVTFTIIGNSPFISSIF APGTTISALIASEFAESDGLHQSGILLIGLLLVLTFTFVLAARLMLRLEKKNAGN" 7490..8335 /gene="pstA" 7490..8335 /gene="pstA" /note="ORF_ID:blrl1093" /codon_start=1	gene
CDS	8332..9153 /gene="pstB" 8332..9153 /gene="pstB" /note="ORF_ID:blrl1094" /codon_start=1 /transl_table=11 /product="phosphate ABC transporter ATP-binding protein" /protein_id="BAC46359.1" /db_xref="GI:27349345" /translation="MSELSVMSAAGLPQAPLLPEAPAKVTVRNLFYGEHALKQ INLTGTRVTAFIFGSGGKSTLIRFNRMVLYPQORATQGLMDQTNILDKLDL NLLRARGVGVFOKPTPTMTIYENIAFGIRLYEIKSKEMDDRVEKALRGALMNEVYK DKLNASGLSLSGQQORLCIARTVAVRPVILLFDEPCSLDPISTAKVELIOLSEN YTIATVTHNMQQARVSDKTAFTAFMYLGLIELBFDDTSKIFTSPSRRTQDYITGREG" 9165..9881 /gene="phoU" 9165..9881 /gene="phoU" /note="ORF_ID:blrl1095" /codon_start=1 /transl_table=11 /product="phosphate transport system regulatory protein" /protein_id="BAC46360.1" /db_xref="GI:27349346" /translation="MGSEHTAKAFDPLQLTRELVAEMGGIAERMIIVDSVDALIRRDV	gene
CDS	13 TCGGGCAGCAGCGGCGGCC 32 Query Match 4.0%; Score 20; DB 1; Length 300900; Best Local Similarity 100.0%; Pred. No. 65; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
	199354 TCGGCGCACGACCGGCGGCC 199373 Db	
RESULT 39	AP005034/c	
LOCUS	AP005034 302300 bp DNA linear BCT 10-MAY-2003	
DEFINITION	Streptomyces avermitilis genomic DNA, complete genome, section 14/30.	
ACCESSION	AP005034 BA000030	
VERSION	AP005034.1 GI:29606779	
KEYWORDS	Streptomyces avermitilis MA-4680	
SOURCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycetaceae; Streptomycetes.	
ORGANISM	Streptomyces avermitilis	
REFERENCE	1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osone, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M. Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites	
AUTHORS	Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)	
TITLE	2 Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis	
JOURNAL		
MEDLINE	21477403	
PUBMED	11572948	
REFERENCE		
AUTHORS		
TITLE		

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
gene  
CDS  
CDS  
CDS  
CDS  
CDS  
CDS

Nat. Biotechnol. 21 (5), 526-531 (2003)  
22608306  
12692562  
3 (bases 1 to 302300)  
Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,  
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,  
Kushida, N., Director-General of Biotechnology Center, Shiba, T.,  
Sakaki, Y., and Hattori, M.  
Direct Submission  
Submitted (29-MAR-2002) Director-General of Biotechnology Center,  
National Institute of Technology and Evaluation, Biotechnology  
Center, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
(E-mail: bio@nite.go.jp, URL: <http://www.bio.nite.go.jp/>,  
Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)  
This work was done in collaboration with Haruo Ikeda(\*1), Jun  
Ishikawa(\*2), Akiharu Hanamoto(\*3), Chigusa Takahashi(\*3), Mayumi  
Shinose(\*3), Hiroshi Horikawa(\*4), Hidekazu Nakazawa(\*4), Tomomi  
Osonoe(\*4), Norihiro Kushida(\*4), Hisashi Kikuchi(\*4), Tadayoshi  
Shiba(\*5), Yoshiyuki Sakaki(\*6,\*7), Masahira Hattori(\*1,\*7)  
and Satoshi Omura(\*1,\*3).  
Final finishing process and all annotation were done by H. Ikeda  
and J. Ishikawa.  
\*1 Kitasato Institute for Life Sciences, Kitasato University  
\*2 National Institute of Infectious Diseases  
\*3 The Kitasato Institute  
\*4 National Institute of Technology and Evaluation  
\*5 School of Science, Kitasato University  
\*6 Institute of Medical Science, University of Tokyo  
\*7 RIKEN, Genomic Sciences Center  
Following url is also available.  
<http://avermittilis.ls.kitasato-u.ac.jp>.  
Location/Qualifiers  
1. .302300  
/organism="Streptomyces avermiltis MA-4680"  
/mol\_type="genomic DNA"  
/strain="MA-4680"  
/db\_xref="taxon:227882"  
/notes="This strain is also named as strain: ATCC 31267,  
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces  
avermectinius"  
/gene="fdxN"  
complement (142. .462)  
/gene="fdxN"  
complement (142. .462)  
/gene="fdxN"  
/note="SAV3129"  
/codon\_start=1  
/transl\_table=11  
/product="putative ferredoxin"  
/protein\_id="BAC70840.1"  
/db\_xref="GI:29606780"  
/translation="MYVIAEPCVDVKACIECPVDCIYEGSRSLYIHPDRCVDCG  
ACFPCVFEALFYEDDTPPEWKDYKANVEFFDELSPGASKLIERDHPFIALP  
PNQ"  
603. .1976  
/note="SAV3130"  
/codon\_start=1  
/transl\_table=11  
/product="putative acetyltransferase"  
/protein\_id="BAC70841.1"  
/db\_xref="GI:29606781"  
/translation="MEISAGQLEVRITADVGKRVSRWNEDVAGGEKFTDTGVYL  
TSWDNGVLLITRGGESVRIAESLVAGKVPAPARRRGPAAEYELARARAWQP  
VERRGDWGLRASGTRTRANSVLPVGEPLPLDEALTVVRDWAYAGRELPAVLOTAT  
GAEGTQLLCALREARGTREVTAELWIGLAPLADREVSABVLSREAEWLSRYQ  
RKGLNEVALVKGSPSWFATVPCAGADPAAGRCVDRGWAGFAAVEVDPAQRR  
GLATVMTLARRALREBSAAALQVETDNAAALYAGMGFATHAYHYRWASED  
VAGLGGRERGAPEAGVSGAGVDEAGAPSEAGAPSEAGAPSEAGAPSESTAGES  
GALGGSAPRGISTPDDVSSAGESDGEVCGSGSAGSVGGPGSGGSGGSDGVDG  
VGTGRSV"  
1973. .2845  
/note="SAV3131"  
/codon\_start=1  
/transl\_table=11

/product="hypothetical protein"  
/protein\_id="BAC70842.1"  
/db\_xref="GI:29606782"  
/translation="MNGHEAMRPHPEPERAAELRRRFADEARSRPDLISALCLLV  
GAADGALDEGIDAQIDLDRLAGQVPRPGPRPWALAEALLGGRCGRGTTPADY  
QRLESSLHEVIRRRRGLPILLSVWMEVARRAGAPVGVALPGHVVVGFGEPRQOVL  
ADPFGGRVLFGDAELLVAGATGAPLDPDSMLTEADPLDVVLRLINIRAWAARPER  
SDVALVAVDLSLLPSHPARLYDRYQALLVERGDFLTGAMELDAYAEVVEADGSAAD  
RIRHQAAARSMLN"  
complement (2864. .3478)  
/note="SAV3132"  
/codon\_start=1  
/transl\_table=11  
/product="putative two-component system response  
regulator"  
/protein\_id="BAC70843.1"  
/db\_xref="GI:29606783"  
/translation="MSRIKVLLEARDQSMVREALAALGLEPDIIEVVAQVAGDRLVA  
AAREHAVALLDIEMPGATGIEAAQVNRRELPAVKLVLTTFGPGVLSRMEAGVA  
AFLVKDAPAAQAEARVKVLAGERVIDPTLAAALAGGANPLUTDREREVIRNAADGAT  
NAELATALHLSQGTVRNYLSTAIQKLAVRNTEAVRTARDKGWL"  
complement (3489. .4646)  
/note="SAV3133"  
/codon\_start=1  
/transl\_table=11  
/product="putative two-component system sensor kinase"  
/protein\_id="BAC70844.1"  
/db\_xref="GI:29606784"  
/translation="MRTGSPENVREALKAMWIVLWVFLSSPVHDLASGHTTPAAT  
AAGWGLAFAFGVYITLTVFRNMGALNGPAVALVAVGLVAVLCLSGPMLGLFV  
YVSVACGTFPLRVAYITPLTAVAMLLVGLHGQDARNILVLLVGLIFAMTGVRL  
VRTVELRKARATQAQLAANEERLADLHLLGHSLSLTIKLSLGAQMLPDHPEK  
AAQOVADIEQVSQALVDREAVTGVTRRVLGELAGMQVLTAAAGVTADLPABDLD  
GVPESEALAWALREAVTNVVRHSGARCVVELVHRQTLDFVLELSVEDDGGSGG  
STPGNGLTGLTERLEKAGSLDAGVRGRFLVARVPMASARGPVCTAGTP"  
complement (4697. .5455)  
/note="SAV3134"  
/codon\_start=1  
/transl\_table=11  
/product="putative ABC transporter integral membrane  
protein"  
/protein\_id="BAC70845.1"  
/db\_xref="GI:29606785"  
/translation="MTNIFASWGLIKLEIVRALNRKFLFFSVIYPSALFLIAGSAD  
STTKVEGTGLSLPTFFVMSAFGLTAVLMGNSRIJAKERESGWQRLLTLPGRG  
VYLKATSAVAISLPSILIVFVAAAVKNDQWLALGTITWVGSILVFAALGVAI  
GYLATGDAVRPITMIVYFGLSMGLGLWPTTTFPDWLQSTAKWLPTHAYAAALQAIEL  
GNAPQAKDIALAVSFVLPAGGAAWLYRKDTLKA"  
complement (5452. .6360)  
/note="SAV3135"  
/codon\_start=1  
/transl\_table=11  
/product="putative ABC transporter ATP-binding protein"  
/protein\_id="BAC70846.1"  
/db\_xref="GI:29606786"  
/translation="MTGVSVAFDQVTKTVYDVRAVDGLTLALHPGTVALLGPNGAG  
KSTLLDLGLKHADSGVRLFGSPREIAVAGRVAGMLQSGMLDEVTVEGLVKAC  
DLHPKPRVTDVLARAGVSIADRVKNKLSGGQEVRFALATAGTDLTLDDEPTG  
MDVTRHAFWATREQADQRTVLFATHYLEADAIADRVILVHRGLLADGTAABEIK  
AKAGARVFDLSGPIDEALRALPELTISDITIRIQSTDADATVHALYGLGVFP  
RNLEVAGLGLGEQARVITAABESKSK"  
complement (6662. .8941)  
/note="SAV3136"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC70847.1"  
/db\_xref="GI:29606787"  
/translation="MSRETSSSAGPNRGGAAYPGTTPYGTPTMAESDGVDAGRPAA  
KSDERKTETTLTRIRINIPGRPIPPVNRFPVDAEADAPAAAPAPGPVSAIS  
TYDVPASASAPGPAAPAGPAEETSDWFAPRKSGAPKPGCGSTNGAGLPTGS  
TVDVPAASAPGPAAPAGPAEETSDWFAPRKSGAPKPGCGSTNGAGLPTGS  
GAQAGCGAGPGGGPRPGAGPGAGPGPGAGPGGAGSGGPGGARNGPVGPG  
QGGGGGTNGAGLPGATGAGVPVAGHGGGTGSGFVSEAVATTSFPFLPVVGSNDGAEP



395 CCGGGGCTGCCGACCACT 377

REFERENCE AUTHORS TITLE	JOURNAL REFERENCE AUTHORS TITLE
1. H. J. ... 2. ... 3. ... 4. ... 5. ... 6. ... 7. ... 8. ... 9. ... 10. ... 11. ... 12. ... 13. ... 14. ... 15. ... 16. ... 17. ... 18. ... 19. ... 20. ... 21. ... 22. ... 23. ... 24. ... 25. ... 26. ... 27. ... 28. ... 29. ... 30. ... 31. ... 32. ... 33. ... 34. ... 35. ... 36. ... 37. ... 38. ... 39. ... 40. ... 41. ... 42. ... 43. ... 44. ... 45. ... 46. ... 47. ... 48. ... 49. ... 50. ... 51. ... 52. ... 53. ... 54. ... 55. ... 56. ... 57. ... 58. ... 59. ... 60. ... 61. ... 62. ... 63. ... 64. ... 65. ... 66. ... 67. ... 68. ... 69. ... 70. ... 71. ... 72. ... 73. ... 74. ... 75. ... 76. ... 77. ... 78. ... 79. ... 80. ... 81. ... 82. ... 83. ... 84. ... 85. ... 86. ... 87. ... 88. ... 89. ... 90. ... 91. ... 92. ... 93. ... 94. ... 95. ... 96. ... 97. ... 98. ... 99. ... 100. ...	1. ... 2. ... 3. ... 4. ... 5. ... 6. ... 7. ... 8. ... 9. ... 10. ... 11. ... 12. ... 13. ... 14. ... 15. ... 16. ... 17. ... 18. ... 19. ... 20. ... 21. ... 22. ... 23. ... 24. ... 25. ... 26. ... 27. ... 28. ... 29. ... 30. ... 31. ... 32. ... 33. ... 34. ... 35. ... 36. ... 37. ... 38. ... 39. ... 40. ... 41. ... 42. ... 43. ... 44. ... 45. ... 46. ... 47. ... 48. ... 49. ... 50. ... 51. ... 52. ... 53. ... 54. ... 55. ... 56. ... 57. ... 58. ... 59. ... 60. ... 61. ... 62. ... 63. ... 64. ... 65. ... 66. ... 67. ... 68. ... 69. ... 70. ... 71. ... 72. ... 73. ... 74. ... 75. ... 76. ... 77. ... 78. ... 79. ... 80. ... 81. ... 82. ... 83. ... 84. ... 85. ... 86. ... 87. ... 88. ... 89. ... 90. ... 91. ... 92. ... 93. ... 94. ... 95. ... 96. ... 97. ... 98. ... 99. ... 100. ...

Result No.	Query			DB	ID	Description
	Score	Match	Length			
C	1	501	100.0	53789	2	AAV21187
	2	36	7.2	20394	2	AAf24892
	3	36	7.2	43280	2	AAf80413
	4	33	6.6	103599	2	ABx04971
	5	32	6.4	29879	2	AAa46806
C	6	29	5.8	77536	3	AAa14651
	7	29	5.8	113193	3	AAa54645
	8	28	5.6	9222	7	AAf61171
C	9	28	5.6	82746	7	AAf61124
	10	27	5.4	665	4	AAf89965
	11	27	5.4	15872	2	AAf68715
12	27	5.4	15872	3	AAa297283	
13	26	5.2	125401	4	AAf71186	
14	25	5.0	9975	7	AAf61173	
15	25	5.0	50937	3	AAa09469	
16	24	4.8	640	4	AAa89130	
17	24	4.8	3978	3	AAc55785	
18	24	4.8	3978	9	AAfE10204	
19	24	4.8	12249	3	AAc55840	
20	24	4.8	12249	9	AAfE10259	
21	24	4.8	18331	3	AAc55857	
22	24	4.8	18331	9	AAfE10266	
23	24	4.8	103599	4	ABx04971	

XX 20-AUG-1996; 96EP-00810551.  
XX (NOVS ) NOVARTIS AG.  
XX Schupp T, Toupet C, Engel N;  
XX WPI; 1998-169172/15.  
XX P-PSDB; AAW52845, AAW52846, AAW52847, AAW52848, AAW52849, AAW52850.  
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used to  
XX produce rifamycin and rifamycin analogues.  
XX Claim 4; Page 53-102; 205pp; English.  
XX The present sequence represents a Amycolatopsis mediterranei rifamycin  
XX synthesis gene cluster DNA fragment from the present invention. The DNA  
XX fragment comprises a DNA region involved directly or indirectly in the  
XX gene cluster responsible for rifamycin synthesis, including the adjacent  
XX DNA regions to the right and left which, by reason of their function in  
XX connection with rifamycin biosynthesis, qualify as constituents of this  
XX rifamycin gene cluster, and functional fragments, derivatives or  
XX constituents of these. The Amycolatopsis mediterranei rifamycin synthesis  
XX gene cluster DNA fragment can be used for producing rifamycin, rifamycin  
XX analogues or precursors. It can also be used for inactivating or  
XX modifying genes involved in ansamycin or rifamycin biosynthesis. The DNA  
XX can be used for constructing mutant actinomycetes strains from which the  
XX natural rifamycin or ansamycin biosynthesis gene cluster has been partly  
XX or completely deleted. The DNA fragment can be used for assembling a  
XX library of polyketide synthases, which can be used for assembling a  
XX library of polyketides. A hybridisation probe of the invention can be  
XX used for identifying DNA fragments involved in the biosynthesis of  
XX ansamycins  
XX  
SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 U; 0 Other;  
Query Match 100.0%; Score 501; DB 2; Length 53789;  
Best Local Similarity 100.0%; Pred. No. 7.2e-214;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCGTCTTCGCTGCTCGCTGCAATGCGGTCAGCGGTCAGGCCCTGACGGCGGCGAGTGTCTCC 60  
DB 8434 TCGTCTTCGCTGCTCGCTGCAATGCGGTCAGCGGTCAGGCCCTGACGGCGGCGAGTGTCTCC 8493  
QY 61 ATGGCGCTCGCGGCGCGCTGATGCTGATGGGCACCGTTCGACGCTTCGTCGAGTTCTCG 120  
DB 8494 ATGGCGCTCGCGGCGCGCTGATGCTGATGGGCACCGTTCGACGCTTCGTCGAGTTCTCG 8553  
QY 121 CGGCAGCGGGGCTGGCCCCCGACGGCGCTGCAAGCGGTTCCGCGACGGCGGACGGC 180  
DB 8554 CGGCAGCGGGGCTGGCCCCCGACGGCGCTGCAAGCGGTTCCGCGACGGCGGACGGC 8613  
QY 181 ACCGCTGTGTCAGGCGCTGCGGCTGCTCTGTTGAGCGGCTGTCGAGGCTCAGCGT 240  
DB 8614 ACCGCTGTGTCAGGCGCTGCGGCTGCTCTGTTGAGCGGCTGTCGAGGCTCAGCGT 8673  
QY 241 CGCGGCGACAGCTCCTCGCGGTGCTCCGCGGTTCGCGGTCACTCCGACGGCGCTCG 300  
DB 8674 CGCGGCGACAGCTCCTCGCGGTGCTCCGCGGTTCGCGGTCACTCCGACGGCGCTCG 8733  
QY 301 AACGGCTTGACGGCCCCCGAACGGCCCCCTGCCAGCAGCGCGTGTATCCGACGGCACTGCCC 360  
DB 8734 AACGGCTTGACGGCCCCCGAACGGCCCCCTGCCAGCAGCGCGTGTATCCGACGGCACTGCCC 8793  
QY 361 GCCCGCGGACTGTCCATCATCGGACGTGACCGCGTGGAGGCGCACGGCACCGGACGACC 420  
DB 8794 GCCCGCGGACTGTCCATCATCGGACGTGACCGCGTGGAGGCGCGCACGGCACCGGACGACC 8853  
QY 421 CTGGGCGACCCGATCGAGCGCGGCGCTGCTGGCCCTACGCGCCAGAACCGGGAACG 480  
DB 8854 CTGGGCGACCCGATCGAGCGCGGCGCTGCTGGCCCTACGCGCCAGAACCGGGAACG 8913  
QY 481 CCGCTGTGGCTCGGCTCGGTG 501

DB 8914 CCGCTGTGGCTCGGTCGGTG 8934  
RESULT 2  
AAF24892  
ID AAF24892 standard; DNA; 20394 BP.  
XX AAF24892;  
XX  
XX 20-APR-2001 (first entry)  
XX Pimaricin biosynthesis associated polyketide synthase gene.  
XX Polyketide synthase; oxidative modification; metabolite; antibiotic;  
XX anticancer; pimaricin; ss.  
XX Streptomyces natalensis.  
XX Key Location/Qualifiers  
XX 1..20394  
XX CDS /\*tag=a  
XX /product= "polyketide synthase"  
XX WO200077222-A1.  
XX 21-DEC-2000.  
XX 14-JUN-2000; 2000WO-EP006227.  
XX 14-JUN-1999; 99EP-00201893.  
XX (STAM ) DSM NV.  
XX Martin JF, Aparicio JF, Colina AJ;  
XX WPI; 2001-080693/09.  
XX P-PSDB; AAB31558.  
XX New polynucleotides encoding enzymes involved in the biosynthesis of  
XX pimaricin, useful for modifying the biosynthesis of pimaricin and in the  
XX synthesis of new compounds.  
XX Disclosure; Page 53-80; 116pp; English.  
XX The present sequence encodes a polyketide synthase which is associated  
XX with the biosynthesis of pimaricin. The polyketide synthase polypeptide  
XX is useful for the oxidative modification of a methyl group of a suitable  
XX compound, e.g. a bioactive compound including a secondary metabolite,  
XX antibiotics and anticancer agents. Recombinant cells comprising the gene  
XX are useful for the production of pimaricin. The polyketide synthase  
XX polynucleotide may be over expressed in Streptomyces, leading to an  
XX increase in the biosynthesis of pimaricin, as a source of primers for  
XX amplification reaction and as probes  
XX  
SQ Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 U; 0 Other;  
Query Match 7.2%; Score 36; DB 4; Length 20394;  
Best Local Similarity 100.0%; Pred. No. 1.5e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 121 CGGCAGCGCGGCTGCGCCCCCGACGGCGCGCTGCAAG 156  
DB 5320 CGGCAGCGCGGCTGCGCCCCCGACGGCGCGCTGCAAG 5355  
RESULT 3  
AAT80413  
ID AAT80413 standard; DNA; 43280 BP.  
XX AAT80413;  
XX 27-FEB-1998 (first entry)

XX Ty lactone synthase gene cluster.

DE Ty lactone synthase gene cluster; tylG gene; multifunctional protein;

XX polyketide; ty lactone synthesis; antibiotic; tylosin; ss.

XX Streptomyces fradiae.

XX Key Location/Qualifiers

FT 816..14243

FT /tag= a

FT /transl\_except= (pos: 816..818, aa: Met)

FT /note= "ORF1 encodes protein shown in AAW22601"

FT 14351..19945

FT /tag= b

FT /transl\_except= (pos: 14351..14353, aa: Met)

FT /note= "ORF2 encodes protein shown in AAW22602"

FT 20010..31199

FT /tag= c

FT /transl\_except= (pos: 20010..20012, aa: Met)

FT /note= "ORF3 encodes protein shown in AAW22603"

FT 31232..36067

FT /tag= d

FT /note= "ORF4 encodes protein shown in AAW22604"

FT 36249..41774

FT /tag= e

FT /note= "ORF5 encodes protein shown in AAW22605"

FT EP791655-A2.

XX 27-AUG-1997.

XX 19-FEB-1997; 97EP-00301056.

XX 22-FEB-1996; 96US-0012078P.

XX (ELIL ) LILLY & CO ELI.

XX Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;

XX WPI; 1997-418046/39.

XX P-PSDB; AAW22601, AAW22602, AAW22603, AAW22604, AAW22605.

XX DNA encoding Streptomyces fradiae ty lactone synthase domain - for

XX production of tylosin-related polyketide compounds.

XX Claim 2; Page 8-66; 220pp; English.

XX This sequence represents the ty lactone synthase gene cluster of the

XX invention. This sequence is also referred to as the tylG gene, and was

XX isolated from Streptomyces fradiae. This sequence encodes multifunctional

XX proteins which direct the synthesis of the polyketide ty lactones, isolated

XX from Streptomyces fradiae. Ty lactone is the basic building block of the

XX antibiotic tylosin. The DNA sequence can be modified so as to alter the

XX type of carboxylic acids incorporated, the number of carboxylic acids

XX incorporated and/or the post-condensation reactions performed, thereby

XX resulting in novel tylosin-related polyketides

XX Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 U; 0 Other;

XX

Query Match 7.2%; Score 36; DB 2; Length 43280;

Best Local Similarity 100.0%; Pred. No. 1.4e-06;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 CTGGCCCCGACGGCGCTGCAAGCGTTCGCCGAC 168

DB 15074 CTGGCCCCGACGGCGCTGCAAGCGTTCGCCGAC 15109

RESULT 4

ABX04971/c

ID ABX04971 standard; DNA; 103599 BP.

XX

AC ABX04971;

XX 16-JAN-2003 (first entry)

XX S. cinnamomensis monensin type I polyketide synthase gene cluster.

DE Monensin; gene; cluster; polyketide synthase; antibiotic; ds;

XX antihelminthic; insecticide; immunosuppressant; antifungal;

XX antibacterial; polyether; mon BI; mon BII; mon CI; mon CII; mon H;

XX mon RI; mon RII; mon T; mon AIX; mon AX.

XX Streptomyces cinnamomensis.

OS WO200168867-A1.

PN 20-SEP-2001.

PD 30-MAY-2000; 2000WO-GB002072.

XX 28-MAY-1999; 99GB-00012563.

XX (BIOT-) BIOTICA TECHNOLOGY LTD.

XX Leadlay PF, Staunton J, Oliynyk M;

XX WPI; 2001-611393/70.

XX P-PSDB; ABG99854, ABG99855, ABG99856, ABG99857, ABG99858, ABG99859,

XX ABG99860, ABG99861, ABG99862, ABG99863, ABG99864, ABG99865, ABG99866,

XX ABG99867, ABG99868, ABG99869, ABG99870, ABG99871, ABG99872, ABG99873,

XX ABG99874, ABG99875, ABG99876, ABG99877, ABG99878, ABG99879, ABG99880,

XX ABG99881, ABG99882, ABG99883, ABG99884, ABG99885, ABG99886, ABG99887.

XX New DNA sequence encoding polyketide synthase, useful for the production

XX of polyketides such as antibiotic monensin.

XX Claim 1; Page 116-195; 212pp; English.

XX The invention relates to a DNA sequence which is a fully defined sequence

XX of 103551 base pairs appearing as ABX04971, or its variant, that it is

XX not a sequence encoding all or part amino acids 1-920 encoded by mon AI

XX as given in the specification. The DNA is the S. cinnamomensis polyketide

XX antibiotic monensin biosynthetic gene cluster. Also included are a

XX recombinant cloning or expression vector comprising the gene cluster, a

XX transformant host cell which has been transformed to contain the gene

XX cluster (and is capable of expressing a corresponding polypeptide), a

XX hybridization probe derived from the gene cluster (for identification and

XX isolation of the same or analogous gene cluster, e.g. one which binds

XX specifically to a region of the monensin gene cluster selected from mon

XX BI, mon BII, mon CI, mon CII, mon H, mon RI, mon RII, mon T, mon AIX and

XX mon AX), the use of the mon RI gene or variant and a monensin promoter to

XX control expression of a heterologous gene in Streptomyces cinnamomensis,

XX a polypeptide encoded by a portion of the monensin gene cluster

XX (preferably comprising mon BI, mon BII, mon CII or mon AX or their

XX mutants, alleles or variants), an epoxide enzyme encoded by mon CI, a

XX cyclase enzyme encoded by mon CII, producing S. cinnamomensis capable of

XX enhanced levels of production of monensin comprising engineering it to

XX overexpress the mon RI gene and/or its variants, expressing a gene homologous to

XX of the mon RI gene and/or its variants, expressing a gene homologous to

XX S. cinnamomensis comprising transforming S. cinnamomensis with DNA

XX encoding a heterologous gene and expressing the gene under control of the

XX activator gene mon RI or actII/orf4 and 13-propyl erythromycin A. The

XX processes and materials (enzyme systems, nucleic acids and vectors) are

XX useful for preparing polyketides by recombinant synthesis. The

XX polyketides are useful as insecticides, antibiotics, antihelminthics,

XX antifungals, antibacterials or other pharmaceuticals. In particular, the

XX gene is useful for the production of monensin, an antibiotic polyether

XX polyketide. The present sequence represents the monensin gene cluster

XX

Sequence 103599 BP; 13980 A; 37023 C; 37799 G; 14795 T; 0 U; 2 Other;

Query Match 6.6%; Score 33; DB 4; Length 103599;

Best Local Similarity 100.0%; Pred. No. 2.7e-05;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 381 GGAGCTGACGCGGTGGAGGCGCACGCGCACCGG 413  
|||||  
DB 83065 GGAGCTGACGCGGTGGAGGCGCACGCGCACCGG 83033  
|||||

RESULT 5  
AAQ46806  
ID AAQ46806 standard; DNA; 29879 BP.  
XX AAQ46806;  
AC  
XX 27-AUG-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 22-DEC-1993 (first entry)  
XX  
DE eryA region of S. erythraea chromosome.  
XX  
XX Saccaropolyspora erythraea; eryA; biosynthesis; polyketide; module;  
KW Saccaropolyspora erythraea; eryA; biosynthesis; polyketide; module;  
KW erythromycin; condensation; elongation; acyl chain growth;  
KW gene replacement; ss.  
XX  
XX Saccharopolyspora erythraea.  
OS  
XX  
XX  
XX Key Location/Qualifiers  
FH CDS 3..10199  
FT /tag= a  
FT /note= "ORF 1"  
FT 10218..20921  
FT /tag= b  
FT /note= "ORF 2"  
FT 20922..29879  
FT /tag= c  
FT /note= "ORF 3"  
XX  
XX WO9313663-A1.  
XX  
XX 22-JUL-1993.  
XX  
XX 17-JAN-1992; 92WO-US000427.  
XX  
XX 17-JAN-1992; 92WO-US000427.  
XX (ABBO ) ABBOTT LAB.  
XX  
XX Katz L, Donadio S, Mcalpine JB;  
XX  
XX WPI; 1993-242804/30.  
DR P-PSDB; AAR44430, AAR44431, AAR44432.  
XX  
XX Biosynthesis of specific polyketide analogues esp. erythromycin cpds. -  
PT by introducing altered biosynthetic gene-contg. DNA into microorganisms.  
XX  
XX Claim 27; Fig 2; 133pp; English.  
XX  
XX This sentence represents a fragment of the Saccaropolyspora erythraea  
CC genome, designated eryA. The polypeptides encoded by this region are  
CC involved in the biosynthesis of the polyketide segment of erythromycin.  
CC eryA is organised in modules and each module takes care of one  
CC condensation step. The precise succession of elongation steps is dictated  
CC by the genetic order of the modules. This fragment may be specifically  
CC altered such that novel polyketide molecules of desired structure are  
CC produced. Three types of alteration may be produced; those inactivating a  
CC single function in a module which does not arrest acyl chain growth;  
CC those inactivating a single function in a module which does affect chain  
CC growth; and those affecting an entire module. The mutations may be  
CC introduced by gene replacement. (Updated on 25-MAR-2003 to correct PN  
CC field.) (Updated on 27-AUG-2003 to correct OS field.)  
XX  
XX Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 U; 0 Other;  
SQ

Query Match 6.4%; Score 32; DB 2; Length 29879;  
Best Local Similarity 100.0%; Pred. No. 8.9e-05;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 382 GACGTGACGCGGTGGAGGCGCACGCGCACCGG 413  
|||||  
DB 11205 GACGTGACGCGGTGGAGGCGCACGCGCACCGG 11236  
|||||

RESULT 6  
AAA14651/c  
ID AAA14651 standard; DNA; 77536 BP.  
XX  
XX AAA14651;  
AC  
XX  
XX 08-AUG-2000 (first entry)  
DT  
XX  
DE Nucleotide sequence of the FK-520 biosynthetic gene cluster.  
XX  
XX FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;  
KW Streptomyces hygroscopicus var. acromyceticus; immunophilin;  
KW FK-506 binding protein; polyketide compound; transplant rejection;  
KW graft-versus-host disease; uveitis; alopecia universalis;  
KW autoimmune chronic active hepatitis; inflammatory bowel disease;  
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;  
KW neurite outgrowth; nerve regrowth; Parkinson's disease;  
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;  
KW peripheral neuropathy; ss.  
XX  
XX Streptomyces hygroscopicus.  
OS  
XX  
XX Key Location/Qualifiers  
FH CDS complement(412..1836)  
FT /tag= a  
FT /note= "fkBW gene"  
FT complement(2020..3579)  
FT /tag= b  
FT /note= "fkBV gene"  
FT 3969..4496  
FT /tag= c  
FT /note= "fkBR2 gene"  
FT complement(4595..5488)  
FT /tag= d  
FT /note= "fkBR1 gene"  
FT 5601..6818  
FT /tag= e  
FT /note= "fkBE gene"  
FT 6808..8052  
FT /tag= f  
FT /note= "fkBF gene"  
FT 8156..8824  
FT /tag= g  
FT /note= "fkBG gene"  
FT complement(9122..9883)  
FT /tag= h  
FT /note= "fkBH gene"  
FT complement(9894..10994)  
FT /tag= i  
FT /note= "fkBI gene"  
FT complement(10987..11247)  
FT /tag= j  
FT /note= "fkBJ gene"  
FT complement(11244..12092)  
FT /tag= k  
FT /note= "fkBK gene"  
FT complement(12113..13150)  
FT /tag= l  
FT /note= "fkBL gene"  
FT complement(13212..23988)  
FT /tag= m  
FT /note= "fkBC gene"  
FT complement(13452..13662)  
FT /tag= n  
FT /note= "ACP6"  
FT  
FT misc\_feature  
FT  
FT misc\_feature  
FT complement(13761..14394)  
FT

```
FT      / *tag= o
FT      /note= "KR6"
FT      complement(14517. .15294)
FT      / *tag= p
FT      /note= "ER6"
FT      complement(15438. .16587)
FT      / *tag= q
FT      /note= "dehydratase domain (DH) 6"
FT      complement(15887. .17820)
FT      / *tag= r
FT      /note= "acyltransferase domain (AT) 6"
FT      complement(17820. .19053)
FT      / *tag= s
FT      /note= "KS6"
FT      complement(19116. .19326)
FT      / *tag= t
FT      /note= "ACP5"
FT      complement(19464. .20097)
FT      / *tag= u
FT      /note= "KS5"
FT      complement(20241. .21420)
FT      / *tag= v
FT      /note= "DHS"
FT      complement(21420. .22653)
FT      / *tag= w
FT      /note= "ATS"
FT      complement(22653. .23892)
FT      / *tag= x
FT      /note= "KSS"
FT      complement(23992. .46573)
FT      / *tag= y
FT      /note= "fkbB gene"
FT      complement(24163. .24373)
FT      / *tag= z
FT      /note= "ACP4"
FT      complement(24997. .26146)
FT      / *tag= aa
FT      /note= "DH4 (inactive)"
FT      complement(26146. .27430)
FT      / *tag= ab
FT      /note= "AT4"
FT      complement(27430. .28684)
FT      / *tag= ac
FT      /note= "KS4"
FT      complement(28750. .28960)
FT      / *tag= ad
FT      /note= "ACP3"
FT      complement(28750. .28960)
FT      / *tag= ae
FT      /note= "DH2 (inactive)"
FT      complement(29092. .29740)
FT      / *tag= af
FT      /note= "KR3"
FT      complement(29869. .31018)
FT      / *tag= ag
FT      /note= "DH3 (inactive)"
FT      complement(31018. .32185)
FT      / *tag= ah
FT      /note= "AT3"
FT      complement(32185. .33439)
FT      / *tag= ai
FT      /note= "KS3"
FT      complement(33505. .33715)
FT      / *tag= aj
FT      /note= "ACP2"
FT      complement(33823. .34480)
FT      / *tag= ak
FT      /note= "KR2"
FT      complement(34606. .35749)
FT      / *tag= al
FT      complement(35749. .37144)
FT      / *tag= am
FT      /note= "AT2"
```

```
FT      misc_feature      complement(37145. .38296)
FT      / *tag= an
FT      /note= "KS2"
FT      complement(38371. .38581)
FT      / *tag= ao
FT      /note= "ACP1"
FT      complement(38677. .39307)
FT      / *tag= ap
FT      /note= "KR1"
FT      complement(39442. .40609)
FT      / *tag= aq
FT      /note= "DH1"
FT      complement(40609. .41842)
FT      / *tag= ar
FT      /note= "AT1"
FT      complement(41842. .43093)
FT      / *tag= as
FT      /note= "KS of extender module 1 (KS1)"
FT      complement(43144. .43660)
FT      / *tag= at
FT      /note= "ACP of loading domain"
FT      complement(43777. .44629)
FT      / *tag= au
FT      /note= "ER of loading domain"
FT      complement(44974. .46573)
FT      / *tag= av
FT      /note= "CoA ligase of loading domain"
FT      46754. .47788
FT      / *tag= aw
FT      /note= "fkbO gene"
FT      47785. .52272
FT      / *tag= ax
FT      /note= "fkbP gene"
FT      52275. .71465
FT      / *tag= ay
FT      /note= "fkbA gene"
FT      52362. .53576
FT      / *tag= az
FT      /note= "KS7"
FT      53577. .54716
FT      / *tag= bb
FT      /note= "AT7"
FT      54717. .55871
FT      / *tag= ba
FT      /note= "DH7"
FT      56019. .56819
FT      / *tag= bc
FT      /note= "ER7"
FT      56943. .57575
FT      / *tag= bd
FT      /note= "KR7"
FT      57710. .57920
FT      / *tag= be
FT      /note= "ACP7"
FT      57990. .59243
FT      / *tag= bf
FT      /note= "KS8"
FT      59244. .60398
FT      / *tag= bg
FT      /note= "AT8"
FT      60399. .61412
FT      / *tag= bh
FT      /note= "DH8 (inactive)"
FT
```

Query Match 5.8%; Score 29; DB 3; Length 77536;  
Best Local Similarity 100.0%; Pred. No. 0.0017;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 GCGGTCAACTCCGACGGCGTCGAACGG 305  
|||||  
Db 37516 GCGGTCAACTCCGACGGCGTCGAACGG 37488





DE Actinosynnema pretiosum polyketide synthase (PKS) gene #2.  
XX Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;  
KW gene; ds.  
XX  
OS Actinosynnema pretiosum.  
XX  
XX WO2003045312-A2.  
XX  
XX  
XX 05-JUN-2003.  
XX  
XX 21-NOV-2002; 2002WO-US037547.  
XX  
XX 21-NOV-2001; 2001US-0332158P.  
XX  
XX (UNIW ) UNIV WASHINGTON.  
XX  
XX Floss HG, Yu T, Leistner E;  
XX WPI; 2003-493374/46.  
XX  
XX Novel maytansinoid produced by bacterial host cell transformed with  
XX expression vector comprising open reading frame from ansamitocin gene  
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX  
XX Disclosure; Page 53-58; 160pp; English.  
XX  
XX The invention relates to maytansinoid produced by bacterial host cell  
XX transformed with expression vector comprising open reading frame from  
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
XX useful as a potent antitumour agent. The present sequence is A. pretiosum  
XX ansamitocin gene cluster I polyketide synthase (PKS) gene  
XX  
XX Sequence 9222 BP; 784 A; 3437 C; 3858 G; 1143 T; 0 U; 0 Other;  
SQ

Query Match 5.6%; Score 28; DB 7; Length 9222;  
Best Local Similarity 100.0%; Pred. No. 0.0062;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 ACGTTCGTCGAGTTCTCGCGCAGCGCG 130  
|||||  
Db 706 ACGTTCGTCGAGTTCTCGCGCAGCGCG 733  
|||||

RESULT 9  
AAL61224/c  
ID AAL61224 standard; DNA; 82746 BP.  
XX  
XX AAL61224;  
AC  
XX  
XX 22-SEP-2003 (first entry)  
DT  
XX Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.  
DE  
XX Maytansinoid; ansamitocin; antitumour; gene; ds.  
KW  
XX Actinosynnema pretiosum.  
OS  
XX  
XX WO2003045312-A2.  
XX  
XX 05-JUN-2003.  
XX  
XX 21-NOV-2002; 2002WO-US037547.  
XX  
XX 21-NOV-2001; 2001US-0332158P.  
XX  
XX (UNIW ) UNIV WASHINGTON.  
XX  
XX Floss HG, Yu T, Leistner E;  
XX WPI; 2003-493374/46.  
XX  
XX Novel maytansinoid produced by bacterial host cell transformed with

PT expression vector comprising open reading frame from ansamitocin gene  
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX  
XX Claim 7; Page 105-152; 160pp; English.  
PS  
XX  
XX The invention relates to maytansinoid produced by bacterial host cell  
XX transformed with expression vector comprising open reading frame from  
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
XX useful as a potent antitumour agent. The present sequence is A. pretiosum  
XX ansamitocin biosynthetic gene cluster I  
XX  
XX Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;  
SQ

Query Match 5.6%; Score 28; DB 7; Length 82746;  
Best Local Similarity 100.0%; Pred. No. 0.0047;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 ACGTTCGTCGAGTTCTCGCGCAGCGCG 130  
|||||  
Db 22257 ACGTTCGTCGAGTTCTCGCGCAGCGCG 22230  
|||||

RESULT 10  
AAF89965  
ID AAF89965 standard; DNA; 665 BP.  
XX  
XX AAF89965;  
AC  
XX  
XX 06-AUG-2001 (first entry)  
DT  
XX Partial nucleotide sequence of a type I polyketide synthase.  
DE  
XX Metabolic pathway operon; polyketide; polyketide antibiotic;  
XX type I polyketide synthase; ss.  
KW  
XX Streptomyces ambofaciens.  
OS  
XX  
XX Key Location/Qualifiers  
EH 25..645  
FT CDS /\*tag= a  
FT /product= "type I polyketide synthase"  
FT  
XX  
XX WO200104097-A2.  
XX  
XX 07-JUN-2001.  
XX  
XX 27-NOV-2000; 2000WO-FR003311.  
XX  
XX 29-NOV-1999; 99FR-00015032.  
XX 07-JUN-2000; 2000US-0209800P.  
XX  
XX (AVET ) AVENTIS PHARMA SA.  
XX  
XX Jeannin P, Pernodet J, Guerineau M, Simonet P, Courtois S;  
XX Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;  
XX Frostegard A;  
XX  
XX WPI; 2001-374849/39.  
XX P-PSDB; AAB83957.  
XX  
XX Collection of nucleic acids from environmental samples, useful for  
XX identifying e.g. genes encoding polyketide synthases and derived  
XX antibiotics.  
XX  
XX Claim 35; Page 220; 356pp; French.  
XX  
XX The specification describes a method for the preparation of a collection  
XX of nucleic acids from organisms in a soil sample. The method comprises  
XX milling a dried sample to produce microparticles; suspending these in  
XX liquid buffer; extraction of nucleic acids from the microparticle;  
XX passing nucleic acid-containing solution through a molecular sieve;  
XX passing nucleic acid-enriched fractions through an anion exchange  
XX chromatography material; and recovering fractions containing purified

CC nucleic acids. The nucleic acids are sources for sequences that encode  
CC either operons involved in a metabolic pathway (specifically polyketide  
CC synthesis) or polypeptides, particularly for production of therapeutic or  
CC agricultural compounds, especially polyketide antibiotics. AAF89964-78  
CC encode partial type I polyketide synthases, and were isolated using the  
CC method of the invention

XX  
SQ Sequence 665 BP; 80 A; 247 C; 244 G; 94 T; 0 U; 0 Other;  
Query Match 5.4%; Score 27; DB 4; Length 665;  
Best Local Similarity 100.0%; Pred. No. 0.024;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 140 CCGACGGCGCTGCGAGCGGTTCGCG 166  
Db 389 CCGACGGCGCTGCGAGCGGTTCGCG 415

RESULT 11  
AAAT68715  
ID AAAT68715 standard; DNA; 15872 BP.  
XX  
AC AAAT68715;  
XX  
DT 01-SEP-1997 (first entry)  
XX  
DE Streptomyces venezuelae polyketide synthase vep ORF1.  
XX  
KW polyketide synthase; polyhydroxyalkanoate monomer synthase;  
KW polyhydroxybutyrate; biodegradable polymer; vep gene;  
KW metabolic engineering; ss.  
XX  
OS Streptomyces venezuelae.  
XX  
FH Key Location/Qualifiers  
FT CDS 20..13912  
FT /\*tag= a  
FT CDS 14056..14136  
FT /\*tag= b  
FT CDS 14148..15827  
FT /\*tag= c  
FT  
FT  
PN W09722711-A1.  
XX  
XX 26-JUN-1997.  
XX  
XX 18-DEC-1996; 96WO-US020119.  
XX  
XX 19-DEC-1995; 95US-0008847P.  
XX  
XX (MINU ) UNIV MINNESOTA.  
XX  
XX Sherman DH, Williams MD, Xue Y;  
XX WPI; 1997-341701/31.  
XX P-PSDB; AAW19629, AAW19630, AAW00918.  
XX  
XX Expression cassettes for production of poly:hydroxyalkanoate(s) -  
XX provide wide range of biodegradable polymers for medical or industrial  
XX use.  
XX  
XX Claim 54; Fig 23; 91pp; English.

XX  
CC Streptomyces venezuelae vep ORF1 (AAAT68715) comprises the polyketide  
CC synthase (PKS) gene cluster encoding a polyene of 12 carbons (see also  
CC AAW19629-30 and AAW00918). It contains 5 PKS modules, with a 5' loading  
CC module and a 3' end domain. Each of the sequenced modules includes a keto  
CC -ACP, an acyltransferase, a dehydratase, a keto- reductase and an acyl  
CC carrier protein domain. The gene cluster was cloned using a heterologous  
CC hybridisation strategy from a genomic DNA library. A novel expression  
CC cassette encoding the first module from the vep gene cluster and module 7  
CC from the Streptomyces tylop gene cluster has polyhydroxyalkanoate (PHA)  
CC monomer synthase activity and can be used for PHA prodn. in host (esp.

CC insect) cells for use as a biodegradable polymer

XX  
SQ Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 U; 0 Other;  
Query Match 5.4%; Score 27; DB 2; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.016;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 GTTCGTCGAGTTCCTCGCGGCGCGG 131  
Db 3721 GTTCGTCGAGTTCCTCGCGGCGCGG 3747

RESULT 12  
AAZ87283  
ID AAZ87283 standard; DNA; 15872 BP.  
XX  
AC AAZ87283;  
XX  
DT 15-SEP-2003 (revised)  
DT 05-JUN-2000 (first entry)  
XX  
DE S. venezuelae vep ORF 1, SEQ ID NO:1.  
XX  
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;  
KW chronic obstructive pulmonary disease; respiratory inflammation;  
KW hypercholesterolaemia; crop protection agent; ds.  
XX  
OS Streptomyces venezuelae; ATCC15439.  
XX  
FH Key Location/Qualifiers  
FT CDS 20..13912  
FT /\*tag= a  
FT /product= "vep ORF 1 amino acid sequence #1 (AAAY77177)"  
FT 14056..14151  
FT /\*tag= b  
FT /product= "vep ORF 1 amino acid sequence #3 (AAAY77199)"  
FT 14167..15827  
FT /\*tag= c  
FT /product= "vep ORF 1 amino acid sequence #2 (AAAY77178)"  
FT  
FT  
PN W0200000620-A2.  
XX  
XX 06-JAN-2000.  
XX  
XX 25-JUN-1999; 99WO-US014398.  
XX  
XX 26-JUN-1998; 98US-00105537.  
XX  
XX (MINU ) UNIV MINNESOTA.  
XX  
XX Sherman DH, Liu H, Xue Y, Zhao L;  
XX WPI; 2000-160679/14.  
XX P-PSDB; AAY77177, AAY77178, AAY77199.  
XX  
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.  
XX synthesis of methymycin and pikromycin.  
XX  
XX Example 3; Fig 23; 438pp; English.

XX  
CC The invention relates to an isolated and purified nucleic acid segment  
CC comprising a desosamine biosynthetic gene cluster, a fragment or its  
CC biologically active variant, where the nucleic acid sequence is not  
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or  
CC Streptomyces antibioticus. The invention also relates to a macrolide  
CC biosynthetic gene cluster, or fragments thereof. The macrolide  
CC biosynthetic gene cluster encodes proteins which synthesize methymycin,  
CC pikromycin, neomethymycin, narbomycin or a combination of these  
CC compounds. Recombinant or augmented cells comprising the desosamine  
CC and/or macrolide biosynthetic gene clusters are useful for the production

of biologically active macrolides. The macrolide biosynthetic proteins are useful for synthesis of methymycin, pikromycin, neomethymycin and narbomycin. The alternative termination of polyketide synthesis may be useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA) monomers. The compounds produced by the recombinant host cells are useful as biopolymers, e.g., in packaging or biomedical applications, to engineer PHA monomer synthases or to prepare biologically active agents, such as chemotherapeutics, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macrolide-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polyketides in plants. The present sequence represents a Streptomyces venezuelae ATCC 15439 DNA sequence, designated vep ORF 1 in the specification, which actually contains 3 open reading frames, which encode proteins AAY7177-Y7178 and AAY7179. The vep ORF 1 protein is defined in the specification as a PHA monomer synthase. (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 U; 0 Other;

Query Match 5.4%; Score 27; DB 3; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.016;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCAGTCTCGCGCAGCGGG 131  
|||||  
DB 3721 GTTCGTCAGTCTCGCGCAGCGGG 3747

## RESULT 13

AA17186  
ID AA17186 standard; DNA; 125401 BP.

XX AA17186;

XX 29-NOV-2001 (first entry)

DE Streptomyces noursei nystatin PKS gene cluster DNA.

XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
KW antifungal; antibiotic; ds.

XX Streptomyces noursei.

Key Location/Qualifiers  
FH 6337..34771  
FT CDS  
FT /\*tag= a  
FT /product= "NysI complete protein"  
FT CDS 34792..51099  
FT /\*tag= b  
FT /product= "NysJ protein"  
FT CDS 51155..57355  
FT /\*tag= c  
FT /product= "NysK protein"  
FT CDS 57503..58687  
FT /\*tag= d  
FT /product= "NysL protein"  
FT CDS complement(58786..58980)  
FT /\*tag= e  
FT /product= "NysM protein"  
FT /note= "CDS does not include start codon"  
FT CDS complement(59045..60241)  
FT /\*tag= f  
FT /product= "NysN protein"  
FT /note= "CDS does not include start codon"  
FT CDS complement(60238..61296)  
FT /\*tag= g  
FT /product= "NysD2 complete protein"  
FT CDS 120628..121308  
FT /\*tag= h

FT /product= "NysR4 (long) protein"

XX W0200159126-A2.

XX 16-AUG-2001.

XX 08-FEB-2001; 2001WO-GB0000509.

XX 08-FEB-2000; 2000GB-00002840.

XX 10-APR-2000; 2000GB-00008786.

XX 14-APR-2000; 2000GB-00009387.

XX (UNNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.

XX (ALPH-) ALPHARMA AS.

XX (SINV-) SINVENT AS.

XX (DZIE/) DZIEGLEWSKA H.

XX (ZOTC/) ZOTCHEV S B.

XX (SEKU/) SEKUROVA O N.

XX (FJAE/) FJAERVIK E.

XX (BRAU/) BRAUTASET T.

XX (STRO/) STROM A R.

XX (VALL/) VALLA S.

XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX WPI; 2001-557614/62.

DR P-PSDB; AA10143, AA10144, AA10145, AA10146, AA10147, AA10148,

DR AA10149, AA10150.

XX New nystatin polyketide synthase polynucleotides and polypeptides, useful

PT as antibiotics and antifungals.

XX Claim 1; Page 188-254; 266pp; English.

XX The present invention relates to the cloning and sequencing of the gene

CC cluster encoding a modular type I polyketide synthase (PKS) enzyme

CC involved in the biosynthesis of the macrolide antibiotic nystatin. The

CC nystatin PKS is useful as antifungal antibiotics. The present sequence is

CC a Streptomyces noursei nystatin PKS gene cluster DNA

XX Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 U; 0 Other;

Query Match 5.2%; Score 26; DB 4; Length 125401;

Best Local Similarity 100.0%; Pred. No. 0.035;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 GACGGCCGCTGCAAGCGTTCGCCGA 167

DB 51905 GACGGCCGCTGCAAGCGTTCGCCGA 51930

RESULT 14

AA161173

ID AA161173 standard; DNA; 9975 BP.

XX AA161173;

XX 22-SEP-2003 (first entry)

XX Actinosynnema pretiosum polyketide synthase (PKS) gene #4.

XX Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;

XX gene; ds.

XX Actinosynnema pretiosum.

XX W02003045312-A2.

XX 05-JUN-2003.

XX 21-NOV-2002; 2002WO-US037547.

PF

```

XX 21-NOV-2001; 2001US-0332158P.
XX (UNIW ) UNIV WASHINGTON.
XX Floss HG, Yu T, Leistner E;
XX WPI; 2003-493374/46.
XX Novel maytansinoid produced by bacterial host cell transformed with
XX expression vector comprising open reading frame from ansamitocin gene
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX Disclosure; Page 62-67; 160pp; English.
XX The invention relates to maytansinoid produced by bacterial host cell
XX transformed with expression vector comprising open reading frame from
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
XX useful as a potent antitumor agent. The present sequence is A. pretiosum
XX ansamitocin gene cluster I polyketide synthase (PKS) gene
XX Sequence 9975 BP; 947 A; 4102 C; 3795 G; 1131 T; 0 U; 0 Other;
XX
XX Query Match 5.0%; Score 25; DB 7; Length 9975;
XX Best Local Similarity 100.0%; Pred.No. 0.13;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 275 CGGCGTCAACTCCGACGCGCGTC 299
DB 5765 CGGCGTCAACTCCGACGCGCGTC 5789
|||||
|||||

RESULT 15
AA09469
ID AAA09469 standard; DNA; 50937 BP.
XX
XX AAA09469;
XX
XX 06-AUG-2003 (revised)
XX 29-AUG-2000 (first entry)
XX
XX Streptococcus oleandomycin gene cluster.
XX
XX oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
XX PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
XX acyl-transferase; acyl carrier protein; inactivated; polyketide;
XX macrolactone; antibiotic; motilide; erythromycin; ss.
XX Streptomyces antibioticus.
XX
XX Key Location/Qualifiers
XX CDS 152..1426
XX FT /*tag= a
XX FT /label= oleI
XX CDS complement(1528..2637)
XX FT /*tag= b
XX FT /label= oleN2
XX CDS complement(2658..4967)
XX FT /*tag= c
XX FT /label= oler
XX CDS 5772..18224
XX FT /*tag= d
XX FT /label= ORF1
XX FT /product= "8,8a-deoxyoleandolide_synthase_1"
XX CDS 18267..29717
XX FT /*tag= e
XX FT /label= ORF2
XX FT /product= "8,8a-deoxyoleandolide_synthase_2"
XX CDS 29787..40346
XX FT /*tag= f
XX FT /label= ORF3
XX FT /product= "8,8a-deoxyoleandolide_synthase_3"
XX CDS 40625..41830

```

```

FT /*tag= g
FT /label= OleP1
FT 41878..43158
FT /*tag= h
FT /label= OleG1
FT 43163..44443
FT /*tag= i
FT /label= OleG2
FT 44433..45173
FT /*tag= j
FT /label= OleM1
FT 45251..46411
FT /*tag= k
FT /label= OleY
FT 46491..47714
FT /*tag= l
FT /label= OleP
FT complement(47808..49517)
FT /*tag= m
FT /label= OleB
XX
XX W0200026349-A2.
XX
XX 11-MAY-2000.
XX
XX 22-OCT-1999; 99WO-US024478.
XX
XX 29-OCT-1998; 98US-0106100P.
XX 16-FEB-1999; 99US-0120254P.
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Betlach MC, Shah SK, Mcdaniel R, Tang L;
XX WPI; 2000-365602/31.
XX P-PSDB; AAY92707, AAY92708, AAY92709.
XX
XX Recombinant DNA compound encoding oleandolide polyketide synthase for
XX synthesizing polyketides comprising a coding sequence for a domain of a
XX loading module or any one of extender modules.
XX
XX Disclosure; Page 14-26; 86pp; English.
XX
XX This is part of the Streptococcus antibioticus oleandomycin gene cluster.
XX The oleandolide polyketide synthase (PKS), also known as 8,8a-
XX deoxyoleandolide synthase, is encoded by three open reading frames (ORF),
XX designated oleAI, oleAII and oleAIII. The PKS is a type I "modular"
XX enzyme, where each ORF encodes 2 extender modules and the first ORF also
XX encodes the loading module. Each module is composed of at least a
XX ketosynthase (KS), acyl-transferase (AT) and an acyl carrier protein
XX (ACP) domain. The oleandolide PKS loading module contains an inactivated
XX KS, called KS-Q, where Q is the abbreviation for glutamine, present
XX instead of the active site cysteine required for activity. The large
XX multifunctional PKS enzymes catalyze the biosynthesis of polyketide
XX macrolactones through multistep pathways involving decarboxylative
XX condensations between acylthioesters followed by cycles of varying beta-
XX carbon processing activities. The macrolide product of the PKS, 8,8a-
XX deoxyoleandolide, is further modified by epoxidation and glycosylation to
XX yield oleandomycin, an antibacterial polyketide. The invention concerns
XX an isolated recombinant DNA compound, comprising a coding sequence for a
XX domain of loading module or any one of extender modules 1-4 or 1-6,
XX including an oleandolide PKS operably linked to a promoter. Also
XX discussed are recombinant oleandolide PKS in which the module 1 KS domain
XX is inactivated by deletion or other mutation. In particular, the
XX inactivation is mediated by a change in the KS domain that renders it
XX incapable of binding substrate (the KS-1 mutation), rendered by mutation
XX in the codon for the active site cysteine. The oleandolide PKS is useful
XX for synthesizing polyketides, which are useful as antibiotics and
XX motilides. Heterologous expression of oleandolide PKS in host cells such
XX as Streptomyces coelicolor and S. lividans is also made possible.
XX Unmodified oleandolide compounds can be provided to cultures of
XX Saccharopolyspora erythraea and converted to the corresponding
XX derivatives of erythromycins A-D. (Updated on 06-AUG-2003 to correct OS

```

```
CC field.)
XX Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 U; 0 Other;
SQ Best Local Similarity 100.0%; Score 25; DB 3; Length 50937;
Query Match 5.0%; Score 25; DB 3; Length 50937;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 GACGGCGCTGCAGGCGTTCGCCG 166
DB 9582 GACGGCGCTGCAGGCGTTCGCCG 9606

RESULT 16
AA89130
ID AA89130 standard; DNA; 640 BP.
XX
AC AA89130;
XX
DT 19-MAR-2001 (first entry)
XX
DE Polyketide synthase derived sequence 34-183R1AS2.T7 DNA.
XX
KW Polyketide synthase; PKS; cosmid 34-183; FK-520; antibiotic;
KW immunosuppressive; neuroimmunophilin; ds.
XX
OS Streptomyces hygroscopicus.
XX
FH Key Location/Qualifiers
FT CDS 1..639
FT FT /*tag= a
FT FT /partial
FT FT /note= "includes in-frame stop codon at 613. .615"
XX
XX US6150513-A.
XX
XX 21-NOV-2000.
XX
XX 16-SEP-1998; 98US-00154083.
XX
XX 16-SEP-1998; 98US-00154083.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Wu X;
XX
XX WPI; 2001-049104/06.
XX
XX P-PSDB; AAB19897.
XX
XX Novel nucleic acid encoding polyketide synthase useful for producing
XX recombinant polyketide synthase for making novel polyketides for use as
XX antibiotic, immunosuppressant or as neuroimmunophilin.
XX
XX Example; Col 21; 30pp; English.
XX
XX The present sequence is that of a polyketide synthase (PKS)-derived
XX coding sequence designated 34-183R1AS2.T7. The DNA was isolated from
XX cosmid 34-183 of an FK-520-producing strain of Streptomyces hygroscopicus
XX following restriction digestion. It does not comprise a full-length
XX coding sequence. The 213-amino acid translated sequence is given in
XX AAB19897. An isolated nucleic acid comprising a coding sequence for a PKS
XX enzyme identical to, or isolated from, a PKS coding sequence contained
XX within a cosmid selected from cosmids 34-183, 34-122 and 34-126 is
XX claimed. Such nucleic acids, when used alone or in combination with other
XX PKS domain coding sequences, are useful in the construction of
XX recombinant vectors that encode PKS enzymes which can be used to make
XX novel polyketides having applications in medicine, agriculture and animal
XX health. The polyketides may have e.g. antibiotic, immunosuppressive or
XX neuroimmunophilin activities
XX
XX Sequence 640 BP; 92 A; 217 C; 244 G; 87 T; 0 U; 0 Other;
SQ Query Match 4.8%; Score 24; DB 4; Length 640;

Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 AGGCGCTGCTGGCCACCTACGGCC 466
DB 98 AGGCGCTGCTGGCCACCTACGGCC 121

RESULT 17
AAC55785
ID AAC55785 standard; DNA; 3978 BP.
XX
AC AAC55785;
XX
DT 19-JAN-2001 (first entry)
XX
DE Type I polyketide synthase orf8.
XX
KW Mitomycin, biosynthesis; mitosome ring system; antibiotic; anti-cancer;
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW fungicide; pesticide; ds.
XX
OS Streptomyces lavendulae.
XX
XX WO200053737-A2.
XX
XX 14-SEP-2000.
XX
XX 10-MAR-2000; 2000WO-US006394.
XX
XX 12-MAR-1999; 99US-00266965.
XX
XX (MINU ) UNIV MINNESOTA.
XX (SHER/) SHERMAN D H.
XX (MAO/) MAO Y..
XX (VARO/) VAROGLU M.
XX (HEMM/) HE M.
XX (SHEL/) SHELTON P C.
XX
XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX WPI; 2000-601980/57.
XX
XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene
XX cluster useful for cloning mitomycin biosynthetic genes for elucidating
XX the molecular basis of mitosome ring system biosynthesis.
XX
XX Example 1; Page 251-253; 399pp; English.
XX
XX This invention relates to isolated and purified nucleic acid molecules
XX from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
XX natural products that contain a variety of functional groups, including
XX amino benzoquinone and axiridine ring systems. The S. lavendulae
XX mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning
XX 55kb of DNA. The invention includes an expression cassette comprising a
XX mitomycin biosynthetic gene operably linked to a promoter, and host cells
XX transformed with the cassette. The nucleotide, and protein sequences and
XX the transformed host cells of the invention result in antiasthmatic,
XX antiinflammatory, cytostatic, immunomodulatory, and antibiotic
XX activities. The nucleotide sequences are used to elucidate the molecular
XX basis for the biosynthesis of the mitosome ring system, as well as to
XX engineer the biosynthesis of novel natural products, e.g. antibiotics,
XX anti-inflammatory agents, anti-cancer agents, immune-enhancers,
XX immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
XX disease as well as other disease involving respiratory inflammation, or
XX cholesterol-lowering agents or as crop protection agents (e.g. fungicides
XX or insecticides) as well as biopolymers, e.g., in packaging or biomedical
XX applications, or to engineer PHA monomer synthases. Sequences AAC55782-
XX AC55881, AAC55815-C55849 and AAB32485-B32542 represent mitomycin
XX biosynthetic gene cluster DNA sequences and encoded proteins. Sequences
XX AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR
XX primers used in the cloning of the mitomycin biosynthetic genes
```

XX SQ Sequence 3978 BP; 497 A; 1583 C; 1415 G; 483 T; 0 U; 0 Other;

Query Match 4.8%; Score 24; DB 3; Length 3978;  
Best Local Similarity 100.0%; Pred. NO. 0.42;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCGACGGCGCTGCAAG 156  
|||||  
Db 757 CTGGCCCCGACGGCGCTGCAAG 780

## RESULT 18

ADE10204  
ID ADE10204 standard; DNA; 3978 BP.

XX AC ADE10204;

XX DT 29-JAN-2004 (first entry)

XX DE S. lavendulae polyketide synthase orfs gene.

XX KW Mitomycin biosynthetic protein; mitomycin C; antibiotic; MC; apoptosis;  
XX KW tumour hypoxia; cytostatic; anti-tumour agent; cancer; ds.

XX OS Streptomyces lavendulae.

XX PN US2003134398-A1.

XX PD 17-JUL-2003.

XX PF 12-SEP-2001; 2001US-00953348.

XX PR 12-SEP-2001; 2001US-00953348.

XX PA (SHER/) SHERMAN D H.

XX PA (MAOY/) MAO Y.

XX PA (VARO/) VAROGLU M.

XX PA (HEMM/) HE M.

XX PA (SHEL/) SHELTON P.

XX PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon P;

XX WPI; 2003-863498/80.

XX New nucleic acid molecule comprising a sequence having mitomycin  
XX biosynthetic gene cluster, useful for enhancing production of  
XX antibiotics.

XX Example 1; SEQ ID NO 19; 308pp; English.

XX The invention relates to an isolated and purified nucleic acid molecule  
XX comprising a sequence having mitomycin biosynthetic gene cluster, or its  
XX variant or fragment. Also included are an expression cassette comprising  
XX the novel nucleic acid molecule operably linked to a promoter functional  
XX in a host cell), a recombinant bacterial host cell in which at least a  
XX portion of a nucleic acid molecule comprising mitomycin biosynthetic gene  
XX cluster is disrupted (resulting in a recombinant host cell that produces  
XX altered levels of mitomycin relative to a corresponding nonrecombinant  
XX bacterial host cell), introducing exogenous DNA into a refractory  
XX Streptomyces strain, identifying a nucleic acid molecule that is related  
XX to at least a portion of a nucleic acid molecule comprising a mitomycin  
XX gene cluster, preparing a compound or its salt from the recombinant host  
XX cell and a product produced by the recombinant host cell. The nucleic  
XX acid encodes a MitT, MitS, MitR, MitO, MitP, MitQ, MitN, MitM, MitL, and/or  
XX MitK, MitJ, MitI, MitH, MitG, MitF, MitE, MitD, MitC, MitA and/or  
XX MmCA-MmcyI. The nucleic acid is useful for enhancing production of  
XX mitomycin antibiotics, which induce apoptosis and hence are useful as  
XX anti-tumour (via tumour hypoxia) agents and are useful in treating  
XX cancer. The gene cluster was isolated from Streptomyces lavendulae. The  
XX present sequence is a gene encoding a non-mitomycin biosynthetic protein  
XX present in the isolated DNA which includes the cluster.

SQ Sequence 3978 BP; 497 A; 1583 C; 1415 G; 483 T; 0 U; 0 Other;

Query Match 4.8%; Score 24; DB 9; Length 3978;  
Best Local Similarity 100.0%; Pred. NO. 0.42;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCGACGGCGCTGCAAG 156  
|||||  
Db 757 CTGGCCCCGACGGCGCTGCAAG 780

## RESULT 19

AAC55840

ID AAC55840 standard; DNA; 12249 BP.

XX AC AAC55840;

XX DT 19-JAN-2001 (first entry)

XX DE Complete Mitomycin ORF 1-9 nucleotide sequence.

XX KW Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;  
XX KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
XX KW chronic obstructive pulmonary disease; respiratory inflammation;  
XX KW fungicide; pesticide; ds.

XX OS Streptomyces lavendulae.

XX PN WO200053737-A2.

XX PD 14-SEP-2000.

XX PF 10-MAR-2000; 2000WO-US006394.

XX PR 12-MAR-1999; 99US-00266965.

XX PA (MINU ) UNIV MINNESOTA.

XX PA (SHER/) SHERMAN D H.

XX PA (MAOY/) MAO Y.

XX PA (VARO/) VAROGLU M.

XX PA (HEMM/) HE M.

XX PA (SHEL/) SHELTON P C.

XX PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;

XX WPI; 2000-601980/57.

XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
XX cluster useful for cloning mitomycin biosynthetic genes for elucidating  
XX the molecular basis of mitosane ring system biosynthesis.

XX Claim 26; Fig 22; 399pp; English.

XX This invention relates to isolated and purified nucleic acid molecules  
XX from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
XX natural products that contain a variety of functional groups, including  
XX amino benzoquinone and axiridine ring systems. The S. lavendulae  
XX mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning  
XX 55Kb of DNA. The invention includes an expression cassette comprising a  
XX mitomycin biosynthetic gene operably linked to a promoter, and host cells  
XX transformed with the cassette. The nucleotide, and protein sequences and  
XX the transformed host cells of the invention result in antiaesthetic,  
XX antiinflammatory, cytostatic, immunomodulatory, and antibiotic  
XX activities. The nucleotide sequences are used to elucidate the molecular  
XX basis for the biosynthesis of the mitosane ring system, as well as to  
XX engineer the biosynthesis of novel natural products, e.g. antibiotics,  
XX anti-inflammatory agents, anti-cancer agents, immune-enhancers,  
XX immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
XX disease as well as other disease involving respiratory inflammation, or  
XX cholesterol-lowering agents or as crop protection agents (e.g. fungicides  
XX or insecticides) as well as biopolymers, e.g., in packaging or biomedical  
XX applications, or to engineer PHA monomer syntheses. Sequences AAC55782-  
XX C55981, AAC55815-C55849 and AAB32485-B32542 represent mitomycin

CC biosynthetic gene cluster DNA sequences and encoded proteins. Sequences  
CC AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR  
CC primers used in the cloning of the mitomycin biosynthetic genes  
XX  
SQ Sequence 12249 BP; 1715 A; 4722 C; 4196 G; 1616 T; 0 U; 0 Other;  
Query Match 4.8%; Score 24; DB 3; Length 12249;  
Best Local Similarity 100.0%; Pred. No. 0.36;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 133 CTGGCCCCGACGGCGCTGCAAG 156  
DB 1744 CTGGCCCCGACGGCGCTGCAAG 1767

RESULT 20  
ADE10259  
ID ADE10259 standard; DNA; 12249 BP.  
XX  
AC ADE10259;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE S. lavendulae mitomycin C gene cluster for ORFs 1-9.  
XX  
KW Mitomycin biosynthetic protein; mitomycin C; antibiotic; MC; apoptosis;  
KW tumour hypoxia; cytostatic; anti-tumour agent; cancer; ds.  
XX  
OS Streptomyces lavendulae.  
XX  
FN US2003134398-A1.  
XX  
PD 17-JUL-2003.  
XX  
PF 12-SEP-2001; 2001US-00953348.  
XX  
PR 12-SEP-2001; 2001US-00953348.  
XX  
PA (SHER/) SHERMAN D H.  
PA (MAOY/) MAO Y.  
PA (VARO/) VAROGLU M.  
PA (HEMM/) HE M.  
PA (SHEL/) SHELDON P.  
XX  
PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon P;  
XX  
DR WPI; 2003-863498/80.  
XX  
PT New nucleic acid molecule comprising a sequence having mitomycin  
PT biosynthetic gene cluster, useful for enhancing production of  
PT antibiotics.  
XX  
PS Claim 21; SEQ ID NO 74; 308pp; English.  
XX  
CC The invention relates to an isolated and purified nucleic acid molecule  
CC comprising a sequence having mitomycin biosynthetic gene cluster, or its  
CC variant or fragment. Also included are an expression cassette comprising  
CC the novel nucleic acid molecule (operably linked to a promoter functional  
CC in a host cell), a recombinant bacterial host cell in which at least a  
CC portion of a nucleic acid molecule comprising mitomycin biosynthetic gene  
CC cluster is disrupted (resulting in a recombinant host cell that produces  
CC altered levels of mitomycin relative to a corresponding nonrecombinant  
CC bacterial host cell), introducing exogenous DNA into a refractory  
CC Streptomyces strain, identifying a nucleic acid molecule that is related  
CC to at least a portion of a nucleic acid molecule comprising a mitomycin  
CC gene cluster, preparing a compound or its salt from the recombinant host  
CC cell and a product produced by the recombinant host cell. The nucleic  
CC acid encodes a MitT, MitS, MitR, MitQ, MitP, MitC, MitB, MitL, MitK,  
CC MitK, MitJ, MitI, MitG, MitF, MitE, MitD, MitC, MitB, MitA and/or  
CC MmcA-MmcY. The nucleic acid is useful for enhancing production of  
CC mitomycin antibiotics, which induce apoptosis and hence are useful as  
CC anti-tumour (via tumour hypoxia) agents and are useful in treating  
CC cancer. The gene cluster was isolated from Streptomyces lavendulae. The

CC present sequence is an operon or gene cluster encoding the mitomycin  
CC biosynthetic proteins of the invention.  
XX  
SQ Sequence 12249 BP; 1715 A; 4722 C; 4196 G; 1616 T; 0 U; 0 Other;  
Query Match 4.8%; Score 24; DB 9; Length 12249;  
Best Local Similarity 100.0%; Pred. No. 0.36;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 133 CTGGCCCCGACGGCGCTGCAAG 156  
DB 1744 CTGGCCCCGACGGCGCTGCAAG 1767

RESULT 21  
AAC55857  
ID AAC55857 standard; DNA; 18331 BP.  
XX  
AC AAC55857;  
XX  
DT 19-JAN-2001 (first entry)  
XX  
DE Complete nucleotide sequence of the mitomycin gene cluster.  
XX  
KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;  
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
KW chronic obstructive pulmonary disease; respiratory inflammation;  
KW fungicide; pesticide; ds.  
XX  
OS Streptomyces lavendulae.  
XX  
FN WO200053737-A2.  
XX  
PD 14-SEP-2000.  
XX  
PF 10-MAR-2000; 2000WO-US006394.  
XX  
PR 12-MAR-1999; 99US-00266965.  
XX  
PA (MINU ) UNIV MINNESOTA.  
PA (SHER/) SHERMAN D H.  
PA (MAOY/) MAO Y.  
PA (VARO/) VAROGLU M.  
PA (HEMM/) HE M.  
PA (SHEL/) SHELDON P C.  
XX  
PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;  
XX  
DR WPI; 2000-601980/57.  
XX  
PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
PT cluster useful for cloning mitomycin biosynthetic genes for elucidating  
PT the molecular basis of mitosome ring system biosynthesis.  
XX  
PS Disclosure; Fig 21; 399pp; English.  
XX  
CC This invention relates to isolated and purified nucleic acid molecules  
CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
CC natural products that contain a variety of functional groups, including  
CC amino benzoquinone and axiridine ring systems. The S. lavendulae  
CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning  
CC 55kb of DNA. The invention includes an expression cassette comprising a  
CC mitomycin biosynthetic gene operably linked to a promoter, and host cells  
CC transformed with the cassette. The nucleotide, and protein sequences and  
CC the transformed host cells of the invention result in antiasthmatic,  
CC antiinflammatory, cytostatic, immunomodulatory, and antibiotic  
CC activities. The nucleotide sequences are used to elucidate the molecular  
CC basis for the biosynthesis of the mitosome ring system, as well as to  
CC engineer the biosynthesis of novel natural products, e.g. antibiotics,  
CC anti-inflammatory agents, anti-cancer agents, immune-enhancers,  
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
CC disease as well as other disease involving respiratory inflammation, or  
CC cholesterol-lowering agents or as crop protection agents (e.g. fungicides





CC (preferably comprising mon BI, mon BII, mon AIX or mon AX or their  
 CC mutants, alleles or variants), an epoxidase enzyme encoded by mon CI, a  
 CC cyclase enzyme encoded by mon CII, producing S. cinammonensis capable of  
 CC enhanced levels of production of monensin comprising engineering it to  
 CC overexpress the mon RI gene. S. cinammonensis containing multiple copies  
 CC of the mon RI gene and/or its variants, expressing a gene heterologous to  
 CC S. cinammonensis comprising transforming S. cinammonensis with DNA  
 CC encoding a heterologous gene and expressing the gene under control of the  
 CC activator gene mon RI or actII/orf4 and 13-propyl erythromycin A. The  
 CC processes and materials (enzyme systems, nucleic acids and vectors) are  
 CC useful for preparing polyketides by recombinant synthesis. The  
 CC polyketides are useful as insecticides, antibiotics, antelmintics,  
 CC antifungals, antibacterials or other pharmaceuticals. In particular the  
 CC gene is useful for the production of monensin, an antibiotic polyether  
 CC polyketide. The present sequence represents the monensin gene cluster  
 XX  
 SQ Sequence 103599 BP; 13980 A; 37023 C; 37799 G; 14795 T; 0 U; 2 Other;  
 Query Match 4.8%; Score 24; DB 4; Length 103599;  
 Best Local Similarity 100.0%; Pred. No. 0.28;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 133 CTGGCCCCGACGGCGCTGCAAG 156  
 DB 55357 CTGGCCCCGACGGCGCTGCAAG 55380  
 RESULT 24  
 AAA14666  
 ID AAA14666 standard; DNA; 4674 BP.  
 AC AAA14666;  
 XX  
 XX 08-AUG-2000 (first entry)  
 DT  
 DE Nucleotide sequence of modified FK-520 PKS gene cluster module 8.  
 XX  
 KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;  
 KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;  
 KW transplant rejection; graft-versus-host disease; alopecia universalis;  
 KW autoimmune chronic active hepatitis; inflammatory bowel disease;  
 KW multiple sclerosis; primary biliary cirrhosis; scleroderma;  
 KW neurite outgrowth; nerve regrowth; Parkinson's disease;  
 KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;  
 KW peripheral neuropathy; ss.  
 XX  
 OS Synthetic.  
 OS Streptomyces hygroscopicus.  
 XX  
 XX Key Location/Qualifiers  
 FH 3. .4673  
 FT CDS /\*tag= a  
 FT /note= "no termination codon given"  
 FT  
 XX  
 PN WO200020601-A2.  
 XX  
 XX 13-APR-2000.  
 XX  
 XX 01-OCT-1999; 99WO-US022886.  
 XX  
 XX 02-OCT-1998; 98US-0102748P.  
 PR 11-MAR-1999; 99US-0123810P.  
 PR 17-JUN-1999; 99US-0139650P.  
 XX  
 XX (KOSA-) KOSAN BIOSCIENCES INC.  
 PA  
 XX Reeves C, Chu D, Khosla C, Santi D, Wu K;  
 XX  
 XX WPI: 2000-317716/27.  
 DR P-PSDB; AAY84730.  
 DR  
 XX New isolated polyketide synthase nucleic acid and polyketide compounds,  
 PT useful for treating e.g. transplant rejection, uveitis, multiple

PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or  
 PT peripheral neuropathy.  
 XX  
 XX Example 2; Page 93-96; 126pp; English.  
 XX  
 CC The present sequence represents module 8 of the FK-520 polyketide  
 CC synthase (PKS) gene cluster, containing the acyltransferase (AT) domain  
 CC of module 12 of rapamycin. FK-506 is a potent immunosuppressant, and acts  
 CC through initial formation of an intermediate complex with protein  
 CC immunophilins known as FK-506 binding proteins. The nucleic acids are  
 CC used for producing polyketide compounds. The polyketide compounds can be  
 CC used as immunosuppressants to prevent or treat transplant rejection,  
 CC graft-versus-host disease or uveitis. They can also be used for treating  
 CC e.g. alopecia universalis, autoimmune chronic active hepatitis,  
 CC inflammatory bowel disease, multiple sclerosis, primary biliary  
 CC cirrhosis, or scleroderma. They also have neurotrophic activity and can  
 CC be used to promote neurite outgrowth in NGF-treated PC12 cells and in  
 CC sensory neuronal cultures, and in intact animals, they promote regrowth  
 CC of damaged facial and sciatic nerves, and repair lesioned serotonin and  
 CC dopamine neurons in the brain. They can also be used for treating e.g.  
 CC Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord  
 CC and brain injury, or peripheral neuropathies. They can also be used in  
 CC agricultural and veterinary applications  
 XX  
 SQ Sequence 4674 BP; 704 A; 1873 C; 1464 G; 633 T; 0 U; 0 Other;  
 Query Match 4.6%; Score 23; DB 3; Length 4674;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 283 AACTCGACGGCGCTGCAACGG 305  
 DB 1254 AACTCGACGGCGCTGCAACGG 1276  
 RESULT 25  
 AAA14665  
 ID AAA14665 standard; DNA; 4725 BP.  
 XX  
 AC AAA14665;  
 XX  
 XX 08-AUG-2000 (first entry)  
 DT  
 XX Nucleotide sequence of FK-520 PKS gene cluster module 8.  
 DE  
 XX  
 KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;  
 KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;  
 KW transplant rejection; graft-versus-host disease; alopecia universalis;  
 KW autoimmune chronic active hepatitis; inflammatory bowel disease;  
 KW multiple sclerosis; primary biliary cirrhosis; scleroderma;  
 KW neurite outgrowth; nerve regrowth; Parkinson's disease;  
 KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;  
 KW peripheral neuropathy; ss.  
 XX  
 OS Streptomyces hygroscopicus.  
 XX  
 XX Key Location/Qualifiers  
 FH 3. .4724  
 FT CDS /\*tag= a  
 FT /note= "no termination codon given"  
 FT  
 XX  
 PN WO200020601-A2.  
 XX  
 XX 13-APR-2000.  
 XX  
 XX 01-OCT-1999; 99WO-US022886.  
 XX  
 XX 02-OCT-1998; 98US-0102748P.  
 PR 11-MAR-1999; 99US-0123810P.  
 PR 17-JUN-1999; 99US-0139650P.  
 XX  
 XX (KOSA-) KOSAN BIOSCIENCES INC.  
 PA  
 XX

PI Reeves C, Chu D, Khosla C, Santi D, Wu K;  
 XX WPI: 2000-317716/27.  
 DR P-PSDB; AAY84729.  
 XX  
 XX New isolated polyketide synthase nucleic acid and polyketide compounds,  
 PT useful for treating e.g. transplant rejection, uveitis, multiple  
 PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or  
 PT peripheral neuropathy.  
 XX  
 PS Example 2; Page 90-93; 126pp; English.  
 XX  
 CC The present sequence encodes module 8 of the FK-520 polyketide synthase  
 CC (PKS) gene cluster of strain MA6548. FK-506 is a potent  
 CC immunosuppressant, and acts through initial formation of an intermediate  
 CC complex with protein immunophilins known as FK-506 binding proteins. The  
 CC nucleic acids are used for producing polyketide compounds. The polyketide  
 CC compounds can be used as immunosuppressants to prevent or treat  
 CC transplant rejection, graft-versus-host disease or uveitis. They can also  
 CC be used for treating e.g. alopecia universalis, autoimmune chronic active  
 CC hepatitis, inflammatory bowel disease, multiple sclerosis, primary  
 CC biliary cirrhosis, or scleroderma. They also have neurotrophic activity  
 CC and can be used to promote neurite outgrowth in NGF-treated PC12 cells  
 CC and in sensory neuronal cultures, and in intact animals, they promote  
 CC regrowth of damaged facial and sciatic nerves, and repair lesioned  
 CC serotonin and dopamine neurons in the brain. They can also be used for  
 CC treating e.g. Parkinson's disease, Alzheimer's disease, stroke, traumatic  
 CC spinal cord and brain injury, or peripheral neuropathies. They can also  
 CC be used in agricultural and veterinary applications  
 XX  
 SQ Sequence 4725 BP; 728 A; 2034 C; 1394 G; 569 T; 0 U; 0 Other;  
 Query Match 4.6%; Score 23; DB 3; Length 4725;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 283 AACTCCGACGGCGCTCGAACGG 305  
 |||||  
 DB 1254 AACTCCGACGGCGCTCGAACGG 1276  
 |||||  
 RESULT 26  
 ID AAA14668 standard; DNA; 4737 BP.  
 AC AAA14668;  
 XX  
 XX 08-AUG-2000 (first entry)  
 DT  
 DE Nucleotide sequence of modified FK-520 PKS gene cluster module 8.  
 XX  
 KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;  
 KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;  
 KW transplant rejection; graft-versus-host disease; alopecia universalis;  
 KW autoimmune chronic active hepatitis; inflammatory bowel disease;  
 KW multiple sclerosis; primary biliary cirrhosis; scleroderma;  
 KW neurite outgrowth; nerve regrowth; Parkinson's disease;  
 KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;  
 KW peripheral neuropathy; ss.  
 XX  
 OS Synthetic.  
 OS Streptomyces hygroscopicus.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 3..4736  
 FT /tag= a  
 FT /note= "no termination codon given"  
 XX  
 XX WO200020601-A2.  
 XX  
 XX 13-APR-2000.  
 XX  
 XX 01-OCT-1999; 99WO-US022886.

XX  
 PR 02-OCT-1998; 98US-0102748P.  
 PR 11-MAR-1999; 99US-0123810P.  
 PR 17-JUN-1999; 99US-0139650P.  
 XX  
 XX (KOSA-) KOSAN BIOSCIENCES INC.  
 XX  
 XX Reeves C, Chu D, Khosla C, Santi D, Wu K;  
 DR WPI: 2000-317716/27.  
 DR P-PSDB; AAY84732.  
 XX  
 PT New isolated polyketide synthase nucleic acid and polyketide compounds,  
 PT useful for treating e.g. transplant rejection, uveitis, multiple  
 PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or  
 PT peripheral neuropathy.  
 XX  
 PS Example 2; Page 99-102; 126pp; English.  
 XX  
 CC The present sequence represents module 8 of the FK-520 polyketide  
 CC synthase (PKS) gene cluster, containing the acyltransferase (AT) domain  
 CC of module 12 of rapamycin. FK-506 is a potent immunosuppressant, and acts  
 CC through initial formation of an intermediate complex with protein  
 CC immunophilins known as FK-506 binding proteins. The nucleic acids are  
 CC used for producing polyketide compounds. The polyketide compounds can be  
 CC used as immunosuppressants to prevent or treat transplant rejection,  
 CC graft-versus-host disease or uveitis. They can also be used for treating  
 CC e.g. alopecia universalis, autoimmune chronic active hepatitis,  
 CC inflammatory bowel disease, multiple sclerosis, primary biliary  
 CC cirrhosis, or scleroderma. They also have neurotrophic activity and can  
 CC be used to promote neurite outgrowth in NGF-treated PC12 cells and in  
 CC sensory neuronal cultures, and in intact animals, they promote regrowth  
 CC of damaged facial and sciatic nerves, and repair lesioned serotonin and  
 CC dopamine neurons in the brain. They can also be used for treating e.g.  
 CC Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord  
 CC and brain injury, or peripheral neuropathies. They can also be used in  
 CC agricultural and veterinary applications  
 XX  
 SQ Sequence 4737 BP; 718 A; 1927 C; 1472 G; 620 T; 0 U; 0 Other;  
 Query Match 4.6%; Score 23; DB 3; Length 4737;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 283 AACTCCGACGGCGCTCGAACGG 305  
 |||||  
 DB 1254 AACTCCGACGGCGCTCGAACGG 1276  
 |||||  
 RESULT 27  
 ID AAA14667 standard; DNA; 4767 BP.  
 XX  
 AC AAA14667;  
 XX  
 XX 08-AUG-2000 (first entry)  
 DT  
 DE Nucleotide sequence of modified FK-520 PKS gene cluster module 8.  
 XX  
 KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;  
 KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;  
 KW transplant rejection; graft-versus-host disease; alopecia universalis;  
 KW autoimmune chronic active hepatitis; inflammatory bowel disease;  
 KW multiple sclerosis; primary biliary cirrhosis; scleroderma;  
 KW neurite outgrowth; nerve regrowth; Parkinson's disease;  
 KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;  
 KW peripheral neuropathy; ss.  
 XX  
 OS Synthetic.  
 OS Streptomyces hygroscopicus.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 3..4766



PA (KOSA-) KOSAN BIOSCIENCES INC.  
 XX Reeves C, Chu D, Khosla C, Santi D, Wu K;  
 PI WPI; 2000-317716/27.  
 XX P-PSDB; AAY84733.  
 DR  
 XX New isolated polyketide synthase nucleic acid and polyketide compounds,  
 PT useful for treating e.g. transplant rejection, uveitis, multiple  
 PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or  
 PT peripheral neuropathy.  
 XX  
 XX Example 2; Page 102-105; 126pp; English.  
 PS  
 XX The present sequence represents module 8 of the FK-520 polyketide  
 CC synthase (PKS) gene cluster, containing the acyltransferase (AT) domain  
 CC of module 13 of rapamycin. FK-506 is a potent immunosuppressant, and acts  
 CC through initial formation of an intermediate complex with protein  
 CC immunophilins known as FK-506 binding proteins. The nucleic acids are  
 CC used for producing polyketide compounds. The polyketide compounds can be  
 CC used as immunosuppressants to prevent or treat transplant rejection,  
 CC graft-versus-host disease or uveitis. They can also be used for treating  
 CC e.g. alopecia universalis, autoimmune chronic active hepatitis,  
 CC inflammatory bowel disease, multiple sclerosis, primary biliary  
 CC cirrhosis, or scleroderma. They also have neurotrophic activity and can  
 CC be used to promote neurite outgrowth in NGF-treated PC12 cells and in  
 CC sensory neuronal cultures, and in intact animals, they promote regrowth  
 CC of damaged facial and sciatic nerves, and repair lesioned serotonin and  
 CC dopamine neurons in the brain. They can also be used for treating e.g.  
 CC Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord  
 CC and brain injury, or peripheral neuropathies. They can also be used in  
 CC agricultural and veterinary applications  
 XX  
 XX Sequence 4818 BP; 742 A; 1982 C; 1476 G; 618 T; 0 U; 0 Other;  
 SQ Query Match 4.6%; Score 23; DB 3; Length 4818;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 283 AACTCCGACGCGCGTCAACGG 305  
 DB 1254 AACTCCGACGCGCGTCAACGG 1276  
 RESULT 30  
 ID AAV21186  
 XX AAV21186 standard; DNA; 5676 BP.  
 XX AC AAV21186;  
 XX  
 DT 24-JUL-1998 (first entry)  
 XX  
 DE Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI DNA fragment.  
 XX  
 XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
 KW polyketide synthase; actinomycete; ansamycin; ds.  
 XX  
 XX Amycolatopsis mediterranei.  
 OS  
 XX Key Location/Qualifiers  
 FH 3..5676  
 FT CDS /tag= a  
 FT /product= "polyketide synthase"  
 FT /note= "no stop codon given"  
 XX  
 XX WO9807868-A1.  
 XX  
 XX 26-FEB-1998.  
 PD  
 XX 18-AUG-1997; 97WO-EP004495.  
 XX  
 XX 20-AUG-1996; 96EP-00810551.  
 XX

PA (NOVS ) NOVARTIS AG.  
 XX Schupp T, Toupet C, Engel N;  
 PI WPI; 1998-169172/15.  
 XX P-PSDB; AAW52844.  
 DR  
 XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used to  
 PT produce rifamycin and rifamycin analogues.  
 PT  
 XX Claim 4; Page 38-43; 205pp; English.  
 PS  
 XX The present sequence represents Amycolatopsis mediterranei strain wt3136  
 CC 5.7 kb KpnI fragment DNA, from the present invention. The present  
 CC invention describes a Amycolatopsis mediterranei rifamycin synthesis gene  
 CC cluster DNA fragment comprising a DNA region involved directly or  
 CC indirectly in the gene cluster responsible for rifamycin synthesis,  
 CC including the adjacent DNA regions to the right and left which, by reason  
 CC of their function in connection with rifamycin biosynthesis, qualify as  
 CC constituents of this rifamycin gene cluster, and functional fragments,  
 CC derivatives or constituents of these. The Amycolatopsis mediterranei  
 CC rifamycin synthesis gene cluster DNA fragment can be used for producing  
 CC rifamycin, rifamycin analogues or precursors. It can also be used for  
 CC inactivating or modifying genes involved in ansamycin or rifamycin  
 CC biosynthesis. The DNA can be used for constructing mutant actinomycetes  
 CC strains from which the natural rifamycin or ansamycin biosynthesis gene  
 CC cluster has been partly or completely deleted. The DNA fragment can be  
 CC used for assembling a library of polyketide synthases, which can be used  
 CC for assembling a library of polyketides. A hybridisation probe of the  
 CC invention can be used for identifying DNA fragments involved in the  
 CC biosynthesis of ansamycins  
 XX  
 XX Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 U; 0 Other;  
 SQ Query Match 4.6%; Score 23; DB 2; Length 5676;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 277 GCGGTCACTCCGACGGCGGTC 299  
 DB 4173 GCGGTCACTCCGACGGCGGTC 4195  
 RESULT 31  
 ID AAF88335  
 XX AAF88335 standard; DNA; 7788 BP.  
 XX AC AAF88335;  
 XX  
 DT 28-AUG-2001 (first entry)  
 XX  
 DE S. spinosa DNA fragment encoding ORF18, SEQ ID 41.  
 XX  
 XX Forosamine; trimethylrharnose; polyketide synthase; biosynthesis;  
 KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;  
 KW macrolide; insecticidal; polyketide synthase; ds.  
 XX  
 XX Saccharopolyspora spinosa.  
 OS  
 XX DE19957268-A1.  
 PN  
 XX 08-MAR-2001.  
 PD  
 XX 29-NOV-1999; 99DE-01057268.  
 PF  
 XX 27-AUG-1999; 99DE-01040596.  
 PR  
 XX (FARB ) BAYER AG.  
 PA  
 XX Eberz G, Moehrle V, Froede R, Velten R, Salas JA;  
 PI WPI; 2001-267102/28.  
 XX P-PSDB; AAB70965.  
 DR



PT enzymes comprises all or a portion of one or more genes in the tmbA gene  
 PT cluster of Sorangium cellulosum.

XX Disclosure; SEQ ID NO 14; 73pp; English.

XX The invention describes a recombinant DNA vector that comprises all or a  
 CC portion of one or more genes in the tmbA gene cluster. Also disclosed is  
 CC the host cell that comprises the tmbA genes of Sorangium cellulosum or  
 CC fragments of those genes. The DNA vector is useful in preparing hybrid  
 CC polyketide synthase (PKS) enzymes and the polyketides produced by such  
 CC hybrid enzymes. The gene products of the tmbA cluster can be used to  
 CC synthesize the polyketide tombamycin. This sequence represents the  
 CC Sorangium cellulosum tmbA gene cluster tmbB open reading frame.

XX Sequence 20922 BP; 2820 A; 6094 C; 8853 G; 3151 T; 0 U; 4 Other;

Query Match 4.6%; Score 23; DB 9; Length 20922;  
 Best Local Similarity 100.0%; Pred. No. 0.95;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 420 CCTGGCGACCCGATCGAGCGCG 442  
 |||||  
 Db 1032 CCTGGCGACCCGATCGAGCGCG 1054

## RESULT 34

AAAD17185  
 ID AAAD17185 standard; DNA; 27541 BP.

XX AC AAAD17185;

XX 29-NOV-2001 (first entry)

XX Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.

XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
 KW antifungal; antibiotic; nys2; ds.

XX Streptomyces noursei.

XX Key Location/Qualifiers  
 FH complement (454..1191)  
 FT /\*tag= a  
 FT /product= "NysF protein"  
 FT /note= "CDS does not include start codon"  
 FT complement (1275..3092)  
 FT /\*tag= b  
 FT /product= "NysG protein"  
 FT complement (3070..4824)  
 FT /\*tag= c  
 FT /note= "CDS does not include start codon"  
 FT /product= "NysH protein"  
 FT 5122..6156  
 FT /\*tag= d  
 FT /product= "NysD3 protein"  
 FT 6338..27541  
 FT /\*tag= e  
 FT /product= "NysI partial protein"  
 FT /note= "CDS does not include stop codon"

XX WO200159126-A2.

XX 16-AUG-2001.

XX 08-FEB-2001; 2001WO-GB000509.

XX 08-FEB-2000; 2000GB-00002840.

XX 10-APR-2000; 2000GB-00008786.

XX 14-APR-2000; 2000GB-00009387.

XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.  
 PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.  
 PA (ALPH-) ALPHARMA AS.

PA (SINV-) SINVENT AS.  
 PA (DZIE/) DZIEGLEWSKA H.  
 PA (ZOTC/) ZOTCHEV S B.  
 PA (SEKU/) SEKUROVA O N.  
 PA (FOAE/) FJAEKVIK E.  
 PA (BRAU/) BRAUTASET T.  
 PA (STRO/) STROM A R.  
 PA (VALL/) VALLA S.

PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;  
 FI Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX WPI; 2001-557614/62.

DR P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.

XX New nystatin polyketide synthase polynucleotides and polypeptides, useful  
 PT as antibiotics and antifungals.

XX Claim 2; Page 151-166; 266pp; English.

XX The present invention relates to the cloning and sequencing of the gene  
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme  
 CC involved in the biosynthesis of the macrolide antibiotic nystatin. The  
 CC nystatin PKS is useful as antifungal antibiotics. The present sequence is  
 CC a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster

XX Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 U; 0 Other;

Query Match 4.6%; Score 23; DB 4; Length 27541;

Best Local Similarity 100.0%; Pred. No. 0.92;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 319 AACGGCCGTCCTCCAGCGCGT 341  
 |||||

Db 21872 AACGGCCGTCCTCCAGCGCGT 21894

## RESULT 35

AAA92301  
 ID AAA92301 standard; DNA; 30690 BP.

XX AC AAA92301;

XX 10-JAN-2001 (first entry)

XX S. avermitilis avermectin aglycon synthase DNA avail SEQ ID NO:1.

XX Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;  
 KW multifunctional enzyme; polyketide; avermectin; veterinary drug;  
 KW agrochemical; ds.

XX Streptomyces avermitilis.

XX Key Location/Qualifiers  
 FH 1..11919  
 FT /\*tag= a  
 FT /note= "avermectin aglycon synthase protein"  
 FT 11971..30690  
 FT /\*tag= b  
 FT /note= "avermectin aglycon synthase protein"

XX WO200050605-A1.

XX 31-AUG-2000.

XX 23-FEB-2000; 2000WO-JP001041.

XX 24-FEB-1999; 99JP-00046961.

XX (KITA) KITASATO INST.

XX Omura S, Ikeda H;  
 PI

XX



DR WPI; 2000-565458/52.  
DR P-PSDB; AAB23749, AAB23750.  
XX  
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of  
PT it for the production of avermectin and its derivatives for drug and  
PT agrochemical use.  
PS  
XX Claim 2; Page 66-134; 314pp; Japanese.  
XX  
CC The present sequence represents DNA which encodes avermectin aglycon  
CC synthase proteins. Also described are: (1) polypeptides encoded by all or  
CC part of the DNA; (2) expression vectors containing the DNA; (3) host  
CC cells transformed by the vectors; (4) preparation of the polypeptides by  
CC culture of the transformants; (5) preparation of avermectin aglycon or  
CC its derivatives by culture of transformed avermectin-producing  
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length  
CC containing sense or antisense sequences from the avermectin aglycon  
CC synthase DNA. The enzymes are useful for the production of modified forms  
CC of avermectin and of the intermediates in its biosynthesis, for use as  
CC drugs, veterinary drugs and agrochemicals  
XX  
SQ Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 3; Length 30690;  
Best Local Similarity 100.0%; Pred. No. 0.91;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 418 ACCCTGGCGACCCGATCGAGGC 440  
Db 13015 ACCCTGGCGACCCGATCGAGGC 13037  
|||||  
RESULT 36  
AAH79277  
ID AAH79277 standard; DNA; 30690 BP.  
XX  
AC AAH79277;  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE Streptomyces avermitilis coding sequences SEQ ID NO: 1.  
XX  
KW Avermectin aglycone synthase; AAS; avermectin derivative;  
KW drug production; veterinary drug; pesticide; ds.  
XX  
OS Streptomyces avermitilis.  
XX  
FH Key Location/Qualifiers  
FT CDS 1. .11919  
FT /\*tag= a  
FT /product= "AAG65264"  
FT /partial  
FT CDS 11971. .30690  
FT /\*tag= b  
FT /product= "AAG65265"  
FT  
PN WO200162939-A1.  
XX  
XX 30-AUG-2001.  
XX  
XX 23-FEB-2001; 2001WO-JP001381.  
XX  
XX 24-FEB-2000; 2000JP-00047405.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
PA (KITA ) KITASATO INST.  
XX  
XX Endo H, Yanaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;  
PI  
XX WPI; 2001-582053/65.  
DR P-PSDB; AAG65264, AAG65265.  
XX  
XX New modified avermectin aglycone synthase derived from Streptomyces

PT avermectilis used in production of 22,23-dihydroavermectin B1a used in  
PT drugs and pesticides.  
XX  
PS Example 2; Page 58-123; 257pp; Japanese.  
XX  
CC The present invention relates to the production of modified derivatives  
CC of avermectin aglycone synthase (AAS) derived from Streptomyces  
CC avermitilis. The activity of an acyl carrier protein (ACP), beta-ketoacyl  
CC carrier protein synthase (KS), acyltransferase (AT), beta-ketoacyl  
CC carrier protein reductase (KR), dehydratase (DH), enoyl reductase (ER)  
CC and/or thioesterase (TE) domain may be reduced or suppressed. The process  
CC can be used in the production of drugs, veterinary drugs and pesticides.  
CC The present sequence is a fragment of the S. avermitilis genome  
XX  
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 4; Length 30690;  
Best Local Similarity 100.0%; Pred. No. 0.91;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 418 ACCCTGGCGACCCGATCGAGGC 440  
Db 13015 ACCCTGGCGACCCGATCGAGGC 13037  
|||||  
RESULT 37  
AA92302  
ID AA92302 standard; DNA; 31422 BP.  
XX  
AC AA92302;  
XX  
DT 10-JAN-2001 (first entry)  
XX  
DE S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.  
XX  
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;  
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;  
KW agrochemical; ds.  
XX  
OS Streptomyces avermitilis.  
XX  
FH Key Location/Qualifiers  
FT CDS 1. .14646  
FT /\*tag= a  
FT /note= "avermectin aglycon synthase protein"  
FT 14824. .31422  
FT /\*tag= b  
FT /note= "avermectin aglycon synthase protein"  
FT  
PN WO200050605-A1.  
XX  
XX 31-AUG-2000.  
XX  
XX 23-FEB-2000; 2000WO-JP001041.  
XX  
XX 24-FEB-1999; 99JP-00046961.  
XX  
XX (KITA ) KITASATO INST.  
XX  
XX Omura S, Ikeda H;  
PI  
XX WPI; 2000-565458/52.  
DR P-PSDB; AAB23751, AAB23752.  
XX  
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of  
PT it for the production of avermectin and its derivatives for drug and  
PT agrochemical use.  
XX  
PS Claim 2; Page 134-203; 314pp; Japanese.  
XX  
CC The present sequence represents DNA which encodes avermectin aglycon  
CC synthase proteins. Also described are: (1) polypeptides encoded by all or  
CC part of the DNA; (2) expression vectors containing the DNA; (3) host  
CC cells transformed by the vectors; (4) preparation of the polypeptides by  
CC culture of the transformants; (5) preparation of avermectin aglycon or  
CC its derivatives by culture of transformed avermectin-producing  
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length  
CC containing sense or antisense sequences from the avermectin aglycon  
CC synthase DNA. The enzymes are useful for the production of modified forms  
CC of avermectin and of the intermediates in its biosynthesis, for use as  
CC drugs, veterinary drugs and agrochemicals  
XX  
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 4; Length 30690;  
Best Local Similarity 100.0%; Pred. No. 0.91;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 418 ACCCTGGCGACCCGATCGAGGC 440  
Db 13015 ACCCTGGCGACCCGATCGAGGC 13037  
|||||

CC cells transformed by the vectors; (4) preparation of the polypeptides by  
CC culture of the transformants; (5) preparation of avermectin aglycon or  
CC its derivatives by culture of transformed avermectin-producing  
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length  
CC containing sense or antisense sequences from the avermectin aglycon  
CC synthase DNA. The enzymes are useful for the production of modified forms  
CC of avermectin and of the intermediates in its biosynthesis, for use as  
CC drugs, veterinary drugs and agrochemicals

XX SQ Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 3; Length 31422;  
Best Local Similarity 100.0%; Pred. No. 0.9;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 212 TGGTGGAGCGGCTGTCGAGGCT 234  
Db 5501 TGGTGGAGCGGCTGTCGAGGCT 5523

RESULT 38  
AAH79278  
ID AAH79278 standard; DNA; 31422 BP.  
AC AAH79278;  
XX  
XX  
XX 04-DEC-2001 (first entry)  
XX Streptomyces avermitilis coding sequences SEQ ID NO: 2.  
DE Avermectin aglycone synthase; AAS; avermectin derivative;  
XX drug production; veterinary drug; pesticide; ds.  
KW Streptomyces avermitilis.  
XX  
XX Key Location/Qualifiers  
FH 1..14646  
FT /\*tag= a  
FT /product= "AAG65266"  
FT 14824..31422  
FT /\*tag= b  
FT /product= "AAG65267"  
XX  
XX WO200162939-A1.  
XX  
XX 30-AUG-2001.  
XX  
XX 23-FEB-2001; 2001WO-JP001381.  
XX  
XX 24-FEB-2000; 2000JP-00047405.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX (KITA ) KITASATO INST.  
XX  
XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;  
XX  
XX WPI; 2001-582053/65.  
XX P-PSDB; AAG65266, AAG65267.  
XX  
XX New modified avermectin aglycone synthase derived from Streptomyces  
PT avermitilis used in production of 22,23-dihydroavermectin B1a used in  
PT drugs and pesticides.  
XX  
XX Disclosure; Page 103-149; 257pp; Japanese.  
XX  
XX The present invention relates to the production of modified derivatives  
CC of avermectin aglycone synthase (AAS) derived from Streptomyces  
CC avermitilis. The activity of an acyl carrier protein (ACP), beta-ketoacyl  
CC carrier protein synthase (KS), acyltransferase (AT), beta-ketoacyl  
CC carrier protein reductase (KR), dehydratase (DH), enoyl reductase (ER)  
CC and/or thioesterase (TE) domain may be reduced or suppressed. The process  
CC can be used in the production of drugs, veterinary drugs and pesticides.  
CC The present sequence is a fragment of the S. avermitilis genome

XX SQ Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 4; Length 31422;  
Best Local Similarity 100.0%; Pred. No. 0.9;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 212 TGGTGGAGCGGCTGTCGAGGCT 234  
Db 5501 TGGTGGAGCGGCTGTCGAGGCT 5523

RESULT 39  
AAS17367  
ID AAS17367 standard; DNA; 33529 BP.  
AC AAS17367;  
XX  
XX  
XX 11-SEP-2003 (revised)  
DT 12-MAR-2002 (first entry)  
XX  
XX DNA sequence of S. cellulosum polyketide synthase cosmid, pKOS28-26.  
XX Polyketide synthase; PKS; catalytic domain; ketosynthase domain;  
KW acyl transferase domain; dehydratase domain; ketoreductase domain;  
KW acyl carrier protein domain; pKOS28-26; ds.  
XX  
XX Polyangium cellulosum.  
OS  
XX US6280999-B1.  
PN  
XX 28-AUG-2001.  
PD  
XX 31-AUG-1998; 98US-00144085.  
PF  
XX 22-JAN-1998; 98US-00010809.  
PR  
XX (KOSA-) KOSAN BIOSCIENCE.  
XX Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;  
PI WPI; 2001-606536/69.  
XX  
XX Novel purified, isolated DNA molecule from Sorangium cellulosum having  
PT polyketide open reading frame encoding modules with one or more domains  
PT such as ketosynthase, acyl transferase and acyl carrier protein domains.  
XX  
XX Claim 4; Fig 1; 72pp; English.  
XX  
XX The present invention relates to the isolation of novel Sorangium  
CC cellulosum polyketide synthases (PKS), and the polynucleotide sequences  
CC encoding them. The polyketide synthases include catalytic domains such as  
CC ketosynthase domain, acyl transferase domain, dehydratase domain,  
CC ketoreductase domain and acyl carrier protein domain. A host cell  
CC comprising a PKS ORF (open reading frame) which encodes one or more  
CC PKS domains is useful for producing polyketide synthases from which  
CC polyketides can be produced. The host cells are useful for constructing a  
CC library, where each individual colony of the library represents a colony  
CC with the ability to produce a particular PKS synthase and ultimately a  
CC particular polyketide. The polyketides produced by these colonies can be  
CC used collectively in a panel to represent a library or may be assessed  
CC individually for activity. Colonies in the library are also induced to  
CC produce the relevant synthases and thus to produce the relevant  
CC polyketides to obtain a library of candidate polyketides which can be  
CC screened for binding to desired targets such as receptors, signalling  
CC proteins, etc. The present sequence represents the DNA sequence of cosmid  
CC pKOS28-26 which encodes one or more domains of S. cellulosum PKS. Note:  
CC The present sequence is said to encode the functional domains of S.  
CC cellulosum PKS which correspond to domains or domain subsets of the amino  
CC acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701). (Updated on 11-SEP  
CC -2003 to standardise OS field)  
XX  
XX Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 0 U; 6 Other;  
SQ

Query Match 4.6%; Score 23; DB 5; Length 33529;  
Best Local Similarity 100.0%; Pred. No. 0.9; Indels 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 420 CCTGGCGACCCGATCGAGCGC 442  
DB 16284 CCTGGCGACCCGATCGAGCGC 16306

RESULT 40  
AAT80414  
ID AAT80414 standard; DNA; 44377 BP.  
XX  
AC AAT80414;  
XX  
DT 27-FEB-1998 (first entry)  
XX  
DE Platenolide synthase gene cluster.  
XX  
KW Ty lactone synthase gene cluster; tylG gene; multifunctional protein;  
KW platenolide synthase gene cluster; platenolide production; smg gene;  
KW polyketide; ty lactone synthesis; antibiotic; tylosin; ss.  
XX  
OS Streptomyces ambofaciens.  
XX  
FH Key Location/Qualifiers  
FT CDS 350..14002  
FT /tag= a  
FT /transl\_except= (pos:350..352, aa:Met)  
FT /note= "ORF1 encodes protein shown in AAW22606"  
FT CDS 14045..20036  
FT /tag= b  
FT /note= "ORF2 encodes protein shown in AAW22607"  
FT CDS 20110..31284  
FT /tag= c  
FT /transl\_except= (pos:20111..20113, aa:Met)  
FT /note= "ORF3 encodes protein shown in AAW22608"  
FT CDS 31329..36071  
FT /tag= d  
FT /note= "ORF4 encodes protein shown in AAW22609"  
FT CDS 36155..41830  
FT /tag= e  
FT /note= "ORF5 encodes protein shown in AAW22610"  
XX  
PN EP791655-A2.  
XX  
XX 27-AUG-1997.  
XX  
PF 19-FEB-1997; 97EP-00301056.  
XX  
PR 22-FEB-1996; 96US-0012078P.  
XX  
PA (ELIL ) LILLY & CO ELI.  
XX  
PI Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;  
XX  
XX WPI; 1997-418046/39.  
DR P-PSDB; AAW22606, AAW22607, AAW22608, AAW22609, AAW22610.  
XX  
XX DNA encoding Streptomyces fradiae ty lactone synthase domain - for  
PT production of tylosin-related polyketide compounds.  
XX  
XX Example 2; Page 110-134; 220pp; English.  
XX  
XX This sequence represents the platenolide synthase gene cluster of the  
CC invention. This sequence is referred to as the smg gene, and was  
CC isolated from Streptomyces ambofaciens. This sequence encodes the multi-  
CC functional proteins which direct the synthesis of the polyketide  
CC platenolide. Platenolide is the basic building block of the macrolide  
CC antibiotic spiramycin. This sequence was used along with the tylG gene  
CC (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG  
CC gene is the ty lactone synthase gene cluster of the invention. The tylG

CC sequence was isolated from Streptomyces fradiae, and encodes  
CC multifunctional proteins which direct the synthesis of the polyketide  
CC ty lactone. Ty lactone is the basic building block of the antibiotic  
CC tylosin. The hybrid sequence can be used to transform S. ambofaciens  
CC lacking the smg ORF1 sequence, or S. fradiae lacking the tylG ORF1  
CC sequence, so that they can produce polyketides. The DNA sequence can be  
CC modified so as to alter the type of carboxylic acids incorporated, the  
CC number of carboxylic acids incorporated and/or the post-condensation  
CC reactions performed, thereby resulting in novel tylosin-related  
CC polyketides  
XX  
SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 2; Length 44377;  
Best Local Similarity 100.0%; Pred. No. 0.87;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GAGTTCTCGCGCAGCGCGGCT 134  
DB 4157 GAGTTCTCGCGCAGCGCGGCT 4179

RESULT 41  
AAT78508  
ID AAT78508 standard; DNA; 44377 BP.  
XX  
AC AAT78508;  
XX  
DT 26-FEB-1998 (first entry)  
XX  
DE Platenolide synthase gene cluster.  
XX  
KW Platenolide synthase gene cluster; platenolide production; smg gene;  
KW multi-functional protein; macrolide antibiotic; spiramycin; ss.  
XX  
OS Streptomyces ambofaciens.  
XX  
FH Key Location/Qualifiers  
FT CDS 350..14002  
FT /tag= a  
FT /transl\_except= (pos:350..352, aa:Met)  
FT /note= "ORF1 encodes protein shown in AAW23716"  
FT CDS 14046..20036  
FT /tag= b  
FT /note= "ORF2 encodes protein shown in AAW23717"  
FT CDS 20110..31284  
FT /tag= c  
FT /transl\_except= (pos:20111..20113, aa:Met)  
FT /note= "ORF3 encodes protein shown in AAW23718"  
FT CDS 31329..36071  
FT /tag= d  
FT /note= "ORF4 encodes protein shown in AAW23719"  
FT CDS 36155..41830  
FT /tag= e  
FT /note= "ORF5 encodes protein shown in AAW23720"  
XX  
PN EP791656-A2.  
XX  
XX 27-AUG-1997.  
XX  
PF 19-FEB-1997; 97EP-00301066.  
XX  
PR 22-FEB-1996; 96US-0012050P.  
XX  
XX (ELIL ) LILLY & CO ELI.  
XX  
PI Burgett SG, Kuhstoss SA, Rao RN, Richardson MA, Rosteck PR;  
XX  
XX WPI; 1997-418047/39.  
DR P-PSDB; AAW23716, AAW23717, AAW23718, AAW23719, AAW23720.  
XX  
XX DNA encoding Streptomyces ambofaciens platenolide synthase domain - for  
PT production of spiramycin-related polyketide antibiotics.

XX Claim 9; Page 8-33; 81pp; English.

XX This sequence represents the platenolide synthase gene cluster of the

XX invention. This sequence is referred to as the *srmd* gene, and was

XX isolated from *Streptomyces ambofaciens*. This sequence encodes the multi-

XX functional proteins which direct the synthesis of the polyketide

XX platenolide. Platenolide is the basic building block of the macrolide

XX antibiotic spiramycin. The DNA can be used to produce compounds

XX exhibiting antibiotic activity based on the platenolide structure,

XX including specifically the macrolide antibiotic spiramycin and spiramycin

XX analogues and derivatives. Modifications of the platenolide synthase DNA

XX sequence can be made so as to change the number and type of carboxylic

XX acids incorporated into the growing polyketide chain and to change the

XX kind of post-condensation processing that is conducted

XX

XX Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;

Query Match 4.6%; Score 23; DB 2; Length 44377;

Best Local Similarity 100.0%; Pred.No. 0.87;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GAGTTCGCGGCAGCGGGCT 134

|||||

Db 4157 GAGTTCGCGGCAGCGGGCT 4179

RESULT 42

AAF88312

ID AAF88312 standard; DNA; 50000 BP.

XX

XX AAF88312;

XX

DT 28-AUG-2001 (first entry)

XX

DE S. spinosa DNA fragment SEQ ID 1.

XX

XX Forosamine; trimethylrharnnose; polyketide synthase; biosynthesis;

XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;

XX macrolide; insecticidal; ds.

XX

OS Saccharopolyspora spinosa.

XX

PN DE19957268-A1.

PD 08-MAR-2001.

XX

PF 29-NOV-1999; 99DE-01057268.

XX

PR 27-AUG-1999; 99DE-01040596.

XX

PA (FARB ) BAYER AG.

XX

PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;

XX

WPI; 2001-267102/28.

XX

PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for

XX recombinant production of insecticidal spinosyns and their derivatives.

XX

PS Claim 7; Page 14-31; 354pp; German.

XX

XX This invention describes a novel method nucleic acid (I) and its encoded

XX polypeptide (II) containing at least one region that encodes an enzymatic

XX activity involved in biosynthesis of spinosyns. (I) are used (i) to

XX identify, inactivate or modulate genes involved in the biosynthesis of

XX (II); (ii) to generate a library of polyketide synthases; (iii) for

XX adding forosamine or trimethylrharnnose to a spinosyn or polyketide

XX aglycone; and (iv) for recombinant production of the corresponding

XX enzymes, which are used for production of (II), their precursors or

XX derivatives, including production of transgenic plants that express (II)

XX and thus have increased resistance to insects. (I) are also useful as

XX markers for sequencing of the Saccharopolyspora spinosa genome. (II) are

CC macrolides with insecticidal, but not antibacterial, activity, and can

CC also be used to raise specific antibodies, useful for identifying

CC expression clones in a gene bank. Cells transformed with (I) may produce

CC (II) at significantly increased levels or produce new derivatives of

CC (II). This sequence represents a genomic DNA fragment of the *S. spinosa*

CC genome which contains the coding regions for proteins involved in

CC forosamine, trimethylrharnnose and polyketide synthase biosynthesis

XX

XX Sequence 50000 BP; 8484 A; 16290 C; 16535 G; 8691 T; 0 U; 0 Other;

Query Match 4.6%; Score 23; DB 4; Length 50000;

Best Local Similarity 100.0%; Pred.No. 0.85;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GACGGCACCGGCTGCTCGAGGG 197

|||||

Db 49400 GACGGCACCGGCTGCTCGAGGG 49422

RESULT 43

AAF88316

ID AAF88316 standard; DNA; 50000 BP.

XX

XX AAF88316;

XX

DT 28-AUG-2001 (first entry)

XX

DE S. spinosa DNA fragment SEQ ID 5.

XX

XX Forosamine; trimethylrharnnose; polyketide synthase; biosynthesis;

XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;

XX macrolide; insecticidal; ds.

XX

OS Saccharopolyspora spinosa.

XX

PN DE19957268-A1.

PD 08-MAR-2001.

XX

PF 29-NOV-1999; 99DE-01057268.

XX

PR 27-AUG-1999; 99DE-01040596.

XX

PA (FARB ) BAYER AG.

XX

PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;

XX

WPI; 2001-267102/28.

XX

PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for

XX recombinant production of insecticidal spinosyns and their derivatives.

XX

PS Claim 7; Page 74-91; 354pp; German.

XX

XX This invention describes a novel method nucleic acid (I) and its encoded

XX polypeptide (II) containing at least one region that encodes an enzymatic

XX activity involved in biosynthesis of spinosyns. (I) are used (i) to

XX identify, inactivate or modulate genes involved in the biosynthesis of

XX (II); (ii) to generate a library of polyketide synthases; (iii) for

XX adding forosamine or trimethylrharnnose to a spinosyn or polyketide

XX aglycone; and (iv) for recombinant production of the corresponding

XX enzymes, which are used for production of (II), their precursors or

XX derivatives, including production of transgenic plants that express (II)

XX and thus have increased resistance to insects. (I) are also useful as

XX markers for sequencing of the Saccharopolyspora spinosa genome. (II) are

SQ Sequence 50000 BP; 6944 A; 14312 C; 19170 G; 9574 T; 0 U; 0 Other;  
 Query Match 4.6%; Score 23; DB 4; Length 50000;  
 Best Local Similarity 100.0%; Pred. No. 0.85;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 175 GACGGCACCGCTGTCGCGAGGG 197  
 DB 3776 GACGGCACCGCTGTCGCGAGGG 3798  
 RESULT 44  
 AAD17184  
 ID AAD17184 standard; DNA; 65140 BP.  
 AC  
 AAD17184;  
 DT 29-NOV-2001 (first entry)  
 XX Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.  
 XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
 KW antifungal; antibiotic; nys1; ds.  
 XX Streptomyces noursei.  
 XX Key Location/Qualifiers  
 CDS complement(1..1035)  
 FT /tag= a  
 FT /product= "NysD2 partial protein"  
 FT /note= "CDS does not include stop codon"  
 FT complement(1056..2576)  
 FT /tag= b  
 FT /product= "NysD1 protein"  
 FT 2806..6906  
 FT /tag= c  
 FT /product= "NysA protein"  
 FT 6952..16530  
 FT /tag= d  
 FT /product= "NysB protein"  
 FT 16550..49840  
 FT /tag= e  
 FT /product= "NysC protein"  
 FT 50260..51015  
 FT /tag= f  
 FT /product= "NysE protein"  
 FT 51405..54305  
 FT /tag= g  
 FT /product= "NysR1 protein"  
 FT 54329..57190  
 FT /tag= h  
 FT /product= "NysR2 protein"  
 FT /note= "CDS does not include start codon"  
 FT 57180..59963  
 FT /tag= i  
 FT /product= "NysR3 protein"  
 FT 60415..61047  
 FT /tag= j  
 FT /product= "NysR4 (short) protein"  
 FT /note= "CDS does not include start codon"  
 FT 61736..62497  
 FT /tag= k  
 FT /product= "NysR5 protein"  
 FT /note= "CDS does not include start codon"  
 FT complement(62551..63615)  
 FT /tag= l  
 FT /product= "ORF2 protein"  
 FT /note= "CDS does not include start codon"  
 FT 63765..64961  
 FT /tag= m  
 FT /product= "ORF1 protein"  
 XX WO200159126-A2.

XX 16-AUG-2001.  
 PD 08-FEB-2001; 2001WO-GB000509.  
 XX 08-FEB-2000; 2000GB-00002840.  
 PR 10-APR-2000; 2000GB-00008786.  
 PR 14-APR-2000; 2000GB-00009387.  
 XX (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.  
 PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.  
 PA (ALPH-) ALPHARMA AS.  
 PA (SINV-) SINVENT AS.  
 PA (DZIE/) DZIEGLEWSKA H.  
 PA (ZOTC/) ZOTCHEV S B.  
 PA (SEKU/) SEKUROVA O N.  
 PA (FJAE/) FJAERVIK E.  
 PA (BRAU/) BRAUTASET T.  
 PA (STRO/) STROM A R.  
 PA (VALL/) VALLA S.  
 XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;  
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;  
 PI WPI; 2001-557614/62.  
 DR P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,  
 DR AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.  
 DR New nystatin polyketide synthase polynucleotides and polypeptides, useful  
 XX as antibiotics and antifungals.  
 FT Claim 2; Page 116-151; 266pp; English.  
 XX The present invention relates to the cloning and sequencing of the gene  
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme  
 CC involved in the biosynthesis of the macrolide antibiotic nystatin. The  
 CC nystatin PKS is useful as antifungal antibiotics. The present sequence is  
 CC a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster  
 XX SQ Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 U; 0 Other;  
 Query Match 4.6%; Score 23; DB 4; Length 65140;  
 Best Local Similarity 100.0%; Pred. No. 0.83;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 319 AACGGCCCGTCCGACGAGCGGT 341  
 DB 12523 AACGGCCCGTCCGACGAGCGGT 12545  
 RESULT 45  
 ADC26995  
 ID ADC26995 standard; DNA; 67251 BP.  
 XX  
 AC ADC26995;  
 XX 18-DEC-2003 (first entry)  
 DT  
 XX Sorangium cellulosum tmbA gene cluster.  
 DE recombinant DNA vector; tmbA gene cluster; tmbA; tmbB; tmbC;  
 XX polyketide synthase enzyme; PKS enzyme; hybrid PKS enzyme; tombamycin;  
 KW ds.  
 XX Polyangium cellulorum.  
 OS US2003054547-A1.  
 PN 20-MAR-2003.  
 XX 28-AUG-2001; 2001US-00942025.  
 PF 22-JAN-1998; 98US-00010809.  
 XX

```

PR 31-AUG-1998; 98US-00144085.
PR 15-FEB-2001; 2001US-0271245P.
XX
XX PA (JULI/) JULIEN B.
XX
XX PI Julien B;
XX
XX WPI; 2003-567100/53.
XX
XX New recombinant DNA vector for preparing hybrid polyketide synthase
XX enzymes comprises all or a portion of one or more genes in the tmBA gene
XX cluster of Sorangium cellulosum.
XX
XX Disclosure; SEQ ID NO 1; 73pp; English.
XX
XX The invention describes a recombinant DNA vector that comprises all or a
XX portion of one or more genes in the tmBA gene cluster. Also disclosed is
XX the host cell that comprises the tmBA genes of Sorangium cellulosum or
XX fragments of those genes. The DNA vector is useful in preparing hybrid
XX polyketide synthase (PKS) enzymes and the polyketides produced by such
XX hybrid enzymes. The gene products of the tmBA cluster can be used to
XX synthesize the polyketide tombamycin. This sequence represents the
XX Sorangium cellulosum tmBA gene cluster.
XX
XX Sequence 67251 BP; 9090 A; 19855 C; 27880 G; 10420 T; 0 U; 6 Other;
XX
XX Query Match 4.6%; Score 23; DB 9; Length 67251;
XX Best Local Similarity 100.0%; Pred. No. 0.82;
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 420 CCTGGGCGACCCGATCGAGGCG 442
XX |||||
XX Db 32521 CCTGGGCGACCCGATCGAGGCG 32543
XX
XX Search completed: June 16, 2004, 17:43:54
XX Job time : 250 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:48:58 ; Search time 57.3333 Seconds  
(without alignments)  
4849.367 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934

Perfect score: 501

Sequence: 1 tcgtcttcgtgctgcgct.....cgctgtgctgcggtcggtg 501

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 15

Total number of hits satisfying chosen parameters: 759

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A-COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B-COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/5A-COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/5B-COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS-COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	36	7.2	43280	2	US-08-804-227C-1
2	32	6.4	20235	1	US-07-642-734C-3
3	32	6.4	20235	3	US-08-439-009A-3
4	29	5.8	77536	4	US-09-410-551B-1
5	27	5.4	15872	3	US-09-105-537-1
6	27	5.4	15872	4	US-09-091-609-1
7	27	5.4	15872	3	US-09-091-609-3
8	25	5.0	50937	3	US-09-428-517-1
9	24	4.8	640	3	US-09-154-083-3
10	24	4.8	3978	4	US-09-266-965-19
11	24	4.8	12249	4	US-09-266-965-74
12	24	4.8	18331	4	US-09-266-965-96
13	23	4.6	4674	4	US-09-410-551B-26
14	23	4.6	4725	4	US-09-410-551B-24
15	23	4.6	4737	4	US-09-410-551B-30
16	23	4.6	4767	4	US-09-410-551B-28
17	23	4.6	4818	4	US-09-410-551B-32
18	23	4.6	11219	1	US-07-642-734C-1
19	23	4.6	11219	3	US-08-439-009A-1
20	23	4.6	13987	2	US-08-804-227C-13
21	23	4.6	33529	3	US-09-144-085-3
22	23	4.6	44377	2	US-08-804-227C-7
23	23	4.6	44377	2	US-08-804-198-1
24	23	4.6	80161	3	US-09-036-987A-1
25	23	4.6	80161	3	US-09-370-700-1
26	23	4.6	80161	4	US-09-603-207-1
27	22	4.4	591	3	US-09-010-809-6

c	28	21	4.2	403	3	US-09-060-756-391
c	29	21	4.2	403	4	US-09-670-314-391
	30	21	4.2	13842	3	US-09-105-537-30
	31	21	4.2	36778	3	US-09-105-537-5
	32	21	4.2	38506	3	US-09-320-878-19
	33	21	4.2	38506	4	US-09-141-908-1
	34	21	4.2	38506	4	US-09-657-440-19
	35	20	4.0	258	4	US-09-252-991A-2488
c	36	20	4.0	699	1	US-08-485-721-1
c	37	20	4.0	699	2	US-08-392-935-1
c	38	20	4.0	699	3	US-08-897-236-1
c	39	20	4.0	699	3	US-09-167-874-1
c	40	20	4.0	699	4	US-09-500-253B-1
c	41	20	4.0	699	5	PCT-US93-08326-1
c	42	20	4.0	1180	3	US-08-897-236-10
c	43	20	4.0	1180	3	US-09-167-874-10
c	44	20	4.0	1180	4	US-09-500-253B-10
c	45	20	4.0	1332	4	US-09-252-991A-4685

ALIGNMENTS

RESULT 1  
US-08-804-227C-1  
; Sequence 1, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kunstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 43280 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 816..14234  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14351..19945  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20010..31199  
; FEATURE:

Sequence 391, App  
Sequence 391, App  
Sequence 30, Appl  
Sequence 5, Appl  
Sequence 19, Appl  
Sequence 1, Appl  
Sequence 19, Appl  
Sequence 2488, Ap  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 10, Appl  
Sequence 10, Appl  
Sequence 4685, Ap





```

; OTHER INFORMATION: acyltransferase domain of module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14062..14610
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase of module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14857..15114
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15166..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15172..16569
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16768..17721
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 18379..18921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19149..19398
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19492..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
; US-07-642-734C-3
;
; Query Match 6.4%; Score 32; DB 1; Length 20235;
; Best Local Similarity 100.0%; Pred. No. 9,1e-06;
; Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 382 GACGTCGACGCGGTGGAGCGCACGCCGCGG 413
; |||||
; DB 1006 GACGTCGACGCGGTGGAGCGCACGCCGCGG 1037
; |||||
;
; RESULT 3
; US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, I
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstock
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/DOCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
; STRAIN: NRRL 238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..10722
; OTHER INFORMATION: /codon start= 19
; OTHER INFORMATION: /function= "gene eryA"
; OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
; OTHER INFORMATION: 6-deoxyerythronolide B"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19..4470
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 97..1482
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1693..2670
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3406..3921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4171..4428
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4471..10722
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 4"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4471..5847
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6054..7026
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 4"
; FEATURE:

```

```

; NAME/KEY: misc feature
; LOCATION: 7165..9216
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: dehydratase and enoylreductase domains m"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 9433..9984
; OTHER INFORMATION: /function= "approximate span
; OTHER INFORMATION: beta-ketoreductase of module 4"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10225..10483
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 4"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10723..20235
; OTHER INFORMATION: /codon_start= 10723
; OTHER INFORMATION: /function= "gene =eryA"
; OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
; OTHER INFORMATION: 6-deoxyerythronolide B formation"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10723..15165
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10831..12174
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 12379..13350
; OTHER INFORMATION: /function= "approximatr span of
; OTHER INFORMATION: acyltransferase domain of module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 14062..14610
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase of module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 14857..15114
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 15165..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 15172..16569
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 16768..17721
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 18379..18921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 19149..19398
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
```

```

; LOCATION: 19492..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
; US-08-439-009A-3
Query Match 6.4%; Score 32; DB 3; Length 20235;
Best Local Similarity 100.0%; Pred. No. 9.1e-06; Indels 0; Gaps 0;
Matches 32; Conservative 0; Mismatches 0;
QY 382 GAGTCGACGCGGTGGAGGCGCACGCGCACCGG 413
| | | | | | | | | | | | | | | | | | | |
Db 1006 GAGTCGACGCGGTGGAGGCGCACGCGCACCGG 1037
| | | | | | | | | | | | | | | | | | | |
RESULT 4
US-09-410-551B-1/c
; Sequence 1, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275) ... (71465)
; US-09-410-551B-1
Query Match 5.8%; Score 29; DB 4; Length 77536;
Best Local Similarity 100.0%; Pred. No. 0.00018; Indels 0; Gaps 0;
Matches 29; Conservative 0; Mismatches 0;
QY 277 GGGGTCAACTCGACGCGCGTCGAACGG 305
| | | | | | | | | | | | | | | | | | | |
Db 37516 GGGGTCAACTCGACGCGCGTCGAACGG 37488
| | | | | | | | | | | | | | | | | | | |
RESULT 5
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
```

```
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match      5.4%; Score 27; DB 3; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 131
    |||||
Db 3721 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 3747

RESULT 6
US-09-091-609-1
; Sequence 1, Application US/09091609
; Patent No. 6600029
; GENERAL INFORMATION:
; APPLICANT: SHERMAN, DAVID H.
; APPLICANT: WILLIAMS, MARK D.
; APPLICANT: XUE, YONGQUAN
; TITLE OF INVENTION: METABOLIC ENGINEERING OF
; POLYHYDROXYALKANOATE MONOMER SYNTHASES
; FILE REFERENCE: 600.297US2
; CURRENT APPLICATION NUMBER: US/09/091,609
; CURRENT FILING DATE: 1998-06-19
; EARLIER APPLICATION NUMBER: PCT/US96/20119
; EARLIER FILING DATE: 1996-12-18
; EARLIER APPLICATION NUMBER: 60/008,847
; EARLIER FILING DATE: 1995-12-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (20)...(13909)
US-09-091-609-1

Query Match      5.4%; Score 27; DB 4; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 131
    |||||
Db 3721 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 3747

RESULT 7
US-09-091-609-3
; Sequence 3, Application US/09091609
; Patent No. 6600029
; GENERAL INFORMATION:
; APPLICANT: SHERMAN, DAVID H.
; APPLICANT: WILLIAMS, MARK D.
; APPLICANT: XUE, YONGQUAN
; TITLE OF INVENTION: METABOLIC ENGINEERING OF
; POLYHYDROXYALKANOATE MONOMER SYNTHASES
; FILE REFERENCE: 600.297US2
; CURRENT APPLICATION NUMBER: US/09/091,609
; CURRENT FILING DATE: 1998-06-19
; EARLIER APPLICATION NUMBER: PCT/US96/20119
; EARLIER FILING DATE: 1996-12-18
; EARLIER APPLICATION NUMBER: 60/008,847
; EARLIER FILING DATE: 1995-12-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae

; TYPE: DNA
; NAME/KEY: CDS
; LOCATION: (14148)...(15824)
US-09-091-609-3

Query Match      5.4%; Score 27; DB 4; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 131
    |||||
Db 3721 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 3747

RESULT 8
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

Query Match      5.0%; Score 25; DB 3; Length 50937;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 GAGGGCGCTGCAAGGCGTTCGCGG 166
    |||||
Db 9582 GAGGGCGCTGCAAGGCGTTCGCGG 9606

RESULT 9
US-09-154-083-3
; Sequence 3, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; CONSTRUCTS THEREFOR
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-09-154-083-3

Query Match      4.8%; Score 24; DB 3; Length 640;
Best Local Similarity 100.0%; Pred. No. 0.056;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```
; OTHER INFORMATION: PKS synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4673)
US-09-410-551B-26

Query Match          4.6%; Score 23; DB 4; Length 4674;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      283 AACTCCGACGGCGGTCGAACGG 305
      |||
Db      1254 AACTCCGACGGCGGTCGAACGG 1276

RESULT 14
US-09-410-551B-24
; Sequence 24, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 4725
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; NAME/KEY: CDS
; LOCATION: (3)..(4724)
US-09-410-551B-24

Query Match          4.6%; Score 23; DB 4; Length 4725;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      283 AACTCCGACGGCGGTCGAACGG 305
      |||
Db      1254 AACTCCGACGGCGGTCGAACGG 1276

RESULT 15
US-09-410-551B-30
; Sequence 30, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 4725
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; NAME/KEY: CDS
; LOCATION: (3)..(4724)
US-09-410-551B-24

Query Match          4.6%; Score 23; DB 4; Length 4725;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      283 AACTCCGACGGCGGTCGAACGG 305
      |||
Db      1254 AACTCCGACGGCGGTCGAACGG 1276

RESULT 16
US-09-410-551B-28
; Sequence 28, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 4767
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; NAME/KEY: CDS
; LOCATION: (3)..(4766)
US-09-410-551B-28

Query Match          4.6%; Score 23; DB 4; Length 4767;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      283 AACTCCGACGGCGGTCGAACGG 305
```

```

Db      1254  AACTCCGACGGCGGTGGAACGG 1276
|||||
RESULT 17
US-09-410-551B-32
; Sequence 32, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYMERIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; TITLE OF INVENTION: CONSTRUCTS THEREFOR
; FILE REFERENCE: 30062-20026-00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 4818
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PXS synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4817)
US-09-410-551B-32
Query Match      4.6%; Score 23; DB 4; Length 4818;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      283  AACTCCGACGGCGGTGGAACGG 305
|||||
Db      1254  AACTCCGACGGCGGTGGAACGG 1276
|||||
RESULT 18
US-07-642-734C-1
; Sequence 1, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Edward H. Gorman
; STREET: Abbott Laboratories D377/AF6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```



OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl ACPSynthase of module 2"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 8262..9305  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 2"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 9906..10454  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 2"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10707..10964  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 2"  
US-07-642-734C-1

Query Match 4.6%; Score 23; DB 1; Length 11219;  
Best Local Similarity 100.0%; Pred. No. 0.12; 0; Indels 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 0;

QY 275 CGGCGTCAACTCCGACGCGG 297  
Db 3049 CGGCGTCAACTCCGACGCGG 3071

RESULT 19  
US-08-439-009A-1  
Sequence 1, Application US/08439009A  
Patent No. 6004787  
GENERAL INFORMATION:  
APPLICANT: Donadio, S  
APPLICANT: Katz, L  
APPLICANT: Mcalpine, J B  
TITLE OF INVENTION: Method of Directing Biosynthesis of  
TITLE OF INVENTION: Specific Polyketides  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESS: Steven F. Weinstein  
STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
CITY: Park Rd  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/439,009A  
FILING DATE: 11-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Casuto, Dianne  
REGISTRATION NUMBER: 40,943  
REFERENCE/DOCKET NUMBER: 4952.US.D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847-938-3137  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharopolyspora erythraea

STRAIN: NRRL 2338  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 744..6659  
OTHER INFORMATION: /function= "APPROXIMATE SPAN OF  
OTHER INFORMATION: MODULE 1"  
OTHER INFORMATION: /label= FUNCTION  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 744..11219  
OTHER INFORMATION: /function= "gene= "eryA"  
OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for  
OTHER INFORMATION: 6-deoxyerythronolide B"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 744..1868  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain 1 of module 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1998..2198  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain 1 of module 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2250..3626  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3831..4811  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase 2 domain of module 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 5574..6125  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6369..6626  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain 2 of module 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6678..11219  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 2"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6678..8066  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl ACPSynthase of module 2"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 8262..9305  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 2"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 9906..10454  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 2"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10707..10964  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 2"  
US-08-439-009A-1

Query Match 4.6%; Score 23; DB 3; Length 11219;  
Best Local Similarity 100.0%; Pred. No. 0.12; 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 0; Indels 0;

Thu Jun 17 09:02:24 2004

```

; CURRENT APPLICATION NUMBER: US/09/144,085
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-144-085-3

Query Match      4.6%; Score 23; DB 3; Length 33529;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      420 CCTGGCGACCCGATCGAGCGC 442
Db      16284 CCTGGCGACCCGATCGAGCGC 16306

RESULT 22
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII (DOS) Text only
; CURRENT APPLICATION DATA:
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..13987
US-08-804-227C-13

Query Match      4.6%; Score 23; DB 2; Length 13987;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      112 GAGTTCTCGCGGCGCGCGCT 134
Db      4142 GAGTTCTCGCGGCGCGCGCT 4164

RESULT 21
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; FILE REFERENCE: 30062-20020.20

```

Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 112 GAGTTCTCGCGGCGGCGGCT 134  
Db 4157 GAGTTCTCGCGGCGGCGGCT 4179

RESULT 24  
US-09-036-987A-1  
; Sequence 1, Application US/09036987A  
; Patent No. 6143526  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Treadway, Patti J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; NUMBER OF SEQUENCES: 39  
; TITLE OF INVENTION: Production  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow Agrosciences LLC Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/036,987A  
; FILING DATE: 09-MAR-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R.  
; REGISTRATION NUMBER: 28,479  
; REFERENCE/DOCKET NUMBER: 50,608  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317)337-4816  
; TELEFAX: (317)337-4847  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 80161 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-036-987A-1

Query Match 4.6%; Score 23; DB 3; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 0.098;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 175 GACGGCACCGGCTGGTCGAGGG 197  
Db 24771 GACGGCACCGGCTGGTCGAGGG 24793

RESULT 25  
US-09-370-700-1  
; Sequence 1, Application US/09370700  
; Patent No. 6274350  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C  
; APPLICANT: Crawford, Kathryn P

Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 112 GAGTTCTCGCGGCGGCGGCT 134  
Db 4157 GAGTTCTCGCGGCGGCGGCT 4179

RESULT 23  
US-08-804-198-1  
; Sequence 1, Application US/08804198  
; Patent No. 5945320  
; GENERAL INFORMATION:  
; APPLICANT: Burgett, Stanley G.  
; APPLICANT: Kuntz, Stuart A.  
; APPLICANT: Rao, Nagaraja R.  
; APPLICANT: Richardson, Mark A.  
; APPLICANT: Rostock, Paul R., Jr.  
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PAUL R. CANTRELL 1138  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: Macintosh 7.0  
; SOFTWARE: Microsoft Word 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,198  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CANTRELL, PAUL R.  
; REGISTRATION NUMBER: 36,470  
; REFERENCE/DOCKET NUMBER: P9113  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3885  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-804-198-1

Query Match 4.6%; Score 23; DB 2; Length 44377;

; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patti J  
; APPLICANT: Turner, Jan R  
; APPLICANT: Walldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489 DIV1  
; CURRENT APPLICATION NUMBER: US/09/370,700  
; CURRENT FILING DATE: 1999-08-09  
; EARLIER APPLICATION NUMBER: US 09/36987  
; EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 80161  
; TYPE: DNA  
; ORGANISM: Saccharopolyspora spinosa  
US-09-370-700-1

Query Match 4.6%; Score 23; DB 3; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 0.098;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 175 GACGGCACCGGCTGTCGAGGG 197  
Db 24771 GACGGCACCGGCTGTCGAGGG 24793

RESULT 26  
US-09-603-207-1  
; Sequence 1, Application US/09603207B  
; Patent No. 6521406  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H  
; APPLICANT: Broughton, Mary C  
; APPLICANT: Crawford, Kathryn P  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patti J  
; APPLICANT: Turner, Jan R  
; APPLICANT: Walldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489 DIV1  
; CURRENT APPLICATION NUMBER: US/09/603,207B  
; CURRENT FILING DATE: 2000-06-23  
; EARLIER APPLICATION NUMBER: 09/370,700  
; EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 80161  
; TYPE: DNA  
; ORGANISM: Saccharopolyspora spinosa  
US-09-603-207-1

Query Match 4.6%; Score 23; DB 4; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 0.098;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 175 GACGGCACCGGCTGTCGAGGG 197  
Db 24771 GACGGCACCGGCTGTCGAGGG 24793

RESULT 27  
US-09-010-809-6  
; Sequence 6, Application US/09010809B  
; Patent No. 6090601  
; GENERAL INFORMATION:  
; APPLICANT: Gustafsson, Claes  
; APPLICANT: Betlach, Mary C  
; TITLE OF INVENTION: Epothilone Polyketide Synthases and Encoding DNA  
; TITLE OF INVENTION: Therefor  
; FILE REFERENCE: 30062-2020.00  
; CURRENT APPLICATION NUMBER: US/09/010,809B

; CURRENT FILING DATE: 1998-01-22  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 591  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-010-809-6

Query Match 4.4%; Score 22; DB 3; Length 591;  
Best Local Similarity 100.0%; Pred. No. 0.046;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 386 TCGACGGCGGTGGAGGCGCACGG 407  
Db 233 TCGACGGCGGTGGAGGCGCACGG 254

RESULT 28  
US-09-060-756-391/c  
; Sequence 391, Application US/09060756  
; Patent No. 6183957  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/060,756  
; CURRENT FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 391  
; LENGTH: 403  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (various positions within the sequence)  
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"  
US-09-060-756-391

Query Match 4.2%; Score 21; DB 3; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 393 GGTGGAGGCGCACGGCACCGG 413  
Db 383 GGTGGAGGCGCACGGCACCGG 363

RESULT 29  
US-09-670-314-391/c  
; Sequence 391, Application US/09670314  
; Patent No. 6492506  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/670,314  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 09/060,756  
; PRIOR FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 391
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-670-314-391

Query Match
Best Local Similarity 4.2%; Score 21; DB 4; Length 403;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 GGTGGAGGCGCACGGCACCGG 413
Db 383 GGTGGAGGCGCACGGCACCGG 363

RESULT 30
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

Query Match
Best Local Similarity 4.2%; Score 21; DB 3; Length 13842;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 TGGGCGACCCGATCGAGGCGG 442
Db 1202 TGGGCGACCCGATCGAGGCGG 1222

RESULT 31
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match
Best Local Similarity 4.2%; Score 21; DB 3; Length 36778;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 422 TGGGCGACCCGATCGAGGCGG 442
Db 2943 TGGGCGACCCGATCGAGGCGG 2963

RESULT 32
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match
Best Local Similarity 4.2%; Score 21; DB 3; Length 38506;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 TGGGCGACCCGATCGAGGCGG 442
Db 1085 TGGGCGACCCGATCGAGGCGG 1105

RESULT 33
US-09-141-908-1
; Sequence 1, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
```

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-141-908-1

Query Match 4.2%; Score 21; DB 4; Length 38506;  
Best Local Similarity 100.0%; Pred.No.0.86;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 422 TGGCGACCCGATCGAGCCG 442  
Db 1085 TGGCGACCCGATCGAGCCG 1105

## RESULT 34

US-09-657-440-19  
; Sequence 19, Application US/09657440  
; Patent No. 6509455  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120  
; CURRENT APPLICATION NUMBER: US/09/657,440  
; CURRENT FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: 09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-657-440-19

Query Match 4.2%; Score 21; DB 4; Length 38506;  
Best Local Similarity 100.0%; Pred.No.0.86;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 422 TGGCGACCCGATCGAGCCG 442  
Db 1085 TGGCGACCCGATCGAGCCG 1105

## RESULT 35

US-09-252-991A-2488  
; Sequence 2488, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 2488  
; LENGTH: 258  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-2488

Query Match 4.0%; Score 20; DB 4; Length 258;  
Best Local Similarity 100.0%; Pred.No.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 446 CGTGTGGCCACCTACGGC 465  
Db 131 CGTGTGGCCACCTACGGC 150

## RESULT 36

US-08-485-721-1/c  
; Sequence 1, Application US/08485721  
; Patent No. 5821124  
; GENERAL INFORMATION:  
; APPLICANT: Regeneron Pharmaceuticals, Inc. and  
; APPLICANT: Regents of the University of California  
; TITLE OF INVENTION: Dorsal Tissue Affecting Factor and  
; TITLE OF INVENTION: Compositions  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
; STREET: 777 Old Saw Mill River Road  
; CITY: Tarrytown  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10591  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,721  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/392,935  
; FILING DATE: 02-SEP-1993  
; APPLICATION NUMBER: PCT/US93/08326  
; FILING DATE: 02-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kempler Ph.D., Gail M.  
; REGISTRATION NUMBER: 32,143  
; REFERENCE/DOCKET NUMBER: Reg 132  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 914-347-7000  
; TELEFAX: 914-347-2113  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 699 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..699  
US-08-485-721-1

Query Match 4.0%; Score 20; DB 1; Length 699;  
Best Local Similarity 100.0%; Pred.No.3.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 TGGCCCCGACGCGCGCTGC 153  
Db 346 TGGCCCCGACGCGCGCTGC 327

## RESULT 37

US-08-392-935-1/c  
; Sequence 1, Application US/08392935  
; Patent No. 5843775  
; GENERAL INFORMATION:

APPLICANT: Regeneron Pharmaceuticals, Inc. and  
APPLICANT: Regents of the University of California  
TITLE OF INVENTION: Dorsal Tissue Affecting Factor and  
TITLE OF INVENTION: Compositions  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Regeneron Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill River Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/392,935  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/08326  
FILING DATE: 02-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Kempler Ph.D., Gail M.  
REGISTRATION NUMBER: 32,143  
REFERENCE/DOCKET NUMBER: Reg 132  
TELEPHONE: 914-347-7000  
TELEFAX: 914-347-2113  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 699 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..699  
US-08-392-935-1

Query Match 4.0%; Score 20; DB 2; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TGGCCCCCGACGGCGCTGC 153  
DB 346 TGGCCCCCGACGGCGCTGC 327

RESULT 38  
US-08-897-236-1/c  
Sequence 1, Application US/08897236A  
Patent No. 6075007  
GENERAL INFORMATION:  
APPLICANT: Regeneron Pharmaceuticals, Inc.  
TITLE OF INVENTION: Modified Dorsal Tissue Affecting Factor and Composition  
FILE REFERENCE: REG 133  
CURRENT APPLICATION NUMBER: US/08/897,236A  
CURRENT FILING DATE: 1997-07-17  
NUMBER OF SEQ ID NOS: 23  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 699  
TYPE: DNA  
ORGANISM: human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(696)  
US-08-897-236-1

Query Match 4.0%; Score 20; DB 3; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TGGCCCCCGACGGCGCTGC 153  
DB 346 TGGCCCCCGACGGCGCTGC 327

RESULT 39  
US-09-167-874-1/c  
Sequence 1, Application US/09167874  
Patent No. 6277593  
GENERAL INFORMATION:  
APPLICANT: Valenzuela et al.  
TITLE OF INVENTION: DORSAL TISSUE AFFECTING FACTOR AND COMPOSITIONS  
FILE REFERENCE: REG132-B  
CURRENT APPLICATION NUMBER: US/09/167,874  
CURRENT FILING DATE: 1998-10-07  
EARLIER APPLICATION NUMBER: 08/485,721  
EARLIER FILING DATE: 1995-07-06  
EARLIER APPLICATION NUMBER: 08/392,935  
EARLIER FILING DATE: 1995-09-22  
EARLIER APPLICATION NUMBER: PCT/US93/08326  
EARLIER FILING DATE: 1993-09-02  
EARLIER APPLICATION NUMBER: 07/957,401  
EARLIER FILING DATE: 1992-10-06  
EARLIER APPLICATION NUMBER: 07/950,410  
EARLIER FILING DATE: 1992-09-23  
EARLIER APPLICATION NUMBER: 07/939,954  
EARLIER FILING DATE: 1992-09-03  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 699  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(696)  
US-09-167-874-1

Query Match 4.0%; Score 20; DB 3; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TGGCCCCCGACGGCGCTGC 153  
DB 346 TGGCCCCCGACGGCGCTGC 327

RESULT 40  
US-09-500-253B-1/c  
Sequence 1, Application US/09500253B  
Patent No. 6500640  
GENERAL INFORMATION:  
APPLICANT: Regeneron Pharmaceuticals, Inc.  
TITLE OF INVENTION: Modified Dorsal Tissue Affecting Factor and Composition  
FILE REFERENCE: REG 133-Z  
CURRENT APPLICATION NUMBER: US/09/500,253B  
CURRENT FILING DATE: 2000-02-08  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 699  
TYPE: DNA  
ORGANISM: human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(696)  
US-09-500-253B-1

Query Match 4.0%; Score 20; DB 4; Length 699;



```

; Patent No. 6075007
;
; GENERAL INFORMATION:
;
; APPLICANT: Regeneron Pharmaceuticals, Inc.
; TITLE OF INVENTION: Modified Dorsal Tissue Affecting Factor and Composition
; FILE REFERENCE: REG 133
; CURRENT APPLICATION NUMBER: US/08/897,236A
; CURRENT FILING DATE: 1997-07-17
;
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10

```

Qy 134 TGGCCCCCGACGGCCGCTGC 153  
|||  
Db 766 TGGCCCCCGACGGCCGCTGC 747

RESULT 44  
US-09-500-253B-10/c  
; Sequence 10, Application US/09500253B  
; Patent No. 6500640  
; GENERAL INFORMATION:  
; APPLICANT: Regeneron Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Modified Dorsal Tissue Affecting Factor and Composition  
; FILE REFERENCE: REG 133-Z  
; CURRENT APPLICATION NUMBER: US/09/500,253B  
; CURRENT FILING DATE: 2000-02-08  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 1180  
; TYPE: DNA  
; ORGANISM: Mouse  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (421)..(1161)  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (16)..(16)  
; OTHER INFORMATION: n = a,c,g, or t  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (235)..(235)  
; OTHER INFORMATION: n = a,c,g, or t  
US-09-500-253B-10

Query Match 4.0%; Score 20; DB 4; Length 1180;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TGGCCCCCGACGGCGCTGC 153  
|||  
Db 766 TGGCCCCCGACGGCGCTGC 747

RESULT 45  
US-09-252-991A-4685/c  
; Sequence 4685, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4685  
; LENGTH: 1332  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4685

Query Match 4.0%; Score 20; DB 4; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 TGCTGGCCACCTACGGCCAG 468  
|||  
Db 1283 TGCTGGCCACCTACGGCCAG 1264

Search completed: June 16, 2004, 19:32:38  
Job time : 60.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:56:30 ; Search time 241 Seconds  
(without alignments)  
9488.873 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934

Perfect score: 501

Sequence: 1 tgcgttcgctgctgcgct.....cgctggtcggtcggtcggtg 501

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 2998549 seqs, 2282253817 residues

Word size : 15

Total number of hits satisfying chosen parameters: 3230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*

3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*

10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*

11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*

12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*

13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*

14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*

15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*

16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*

17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*

18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*

19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	6.4	84428	US-10-229-148B-1	Sequence 1, Appli
2	29	5.8	18435	US-10-156-761-412	Sequence 412, App
3	29	5.8	77536	US-09-940-316B-1	Sequence 1, Appli
4	29	5.8	100000	US-10-156-761-15103	Sequence 15103, A
5	29	5.8	9025608	US-10-156-761-1	Sequence 1, Appli
6	27	5.4	14520	US-10-156-761-2885	Sequence 2885, App
7	27	5.4	15872	US-09-861-289-1	Sequence 1, Appli
8	27	5.4	15872	US-09-860-846-1	Sequence 1, Appli
9	27	5.4	15872	US-09-988-384B-1	Sequence 1, Appli
10	27	5.4	15872	US-09-836-821-1	Sequence 1, Appli
11	27	5.4	15872	US-10-271-889-44	Sequence 44, Appl
12	27	5.4	125746	US-10-156-761-15102	Sequence 15102, A
13	26	5.2	10692	US-10-156-761-414	Sequence 414, App
14	26	5.2	64492	US-10-378-083-1	Sequence 1, Appli

15	5.0	50937	10	US-09-808-880-1	Sequence 1, Appli
16	4.8	3978	10	US-09-953-348-19	Sequence 19, Appl
17	4.8	3978	15	US-10-267-255-19	Sequence 19, Appl
18	4.8	5505	15	US-10-156-761-413	Sequence 413, App
19	4.8	5721	15	US-10-156-761-2880	Sequence 2880, Ap
20	4.8	11910	15	US-10-156-761-2879	Sequence 2879, Ap
21	4.8	12249	10	US-09-953-348-74	Sequence 74, Appl
22	4.8	12249	15	US-10-267-255-74	Sequence 74, Appl
23	4.8	18331	10	US-09-953-348-96	Sequence 96, Appl
24	4.8	18331	15	US-10-267-255-96	Sequence 96, Appl
25	4.8	86941	17	US-10-461-194-2	Sequence 2, Appli
26	4.8	125746	15	US-10-156-761-15102	Sequence 15102, A
27	4.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
28	4.6	4674	10	US-09-940-316B-26	Sequence 26, Appl
29	4.6	4725	10	US-09-940-316B-24	Sequence 24, Appl
30	4.6	4737	10	US-09-940-316B-30	Sequence 30, Appl
31	4.6	4767	10	US-09-940-316B-28	Sequence 28, Appl
32	4.6	4818	10	US-09-940-316B-32	Sequence 32, Appl
33	4.6	10056	15	US-10-156-761-411	Sequence 411, App
34	4.6	14055	15	US-10-156-761-2883	Sequence 2883, Ap
35	4.6	14643	15	US-10-156-761-931	Sequence 931, App
36	4.6	18717	15	US-10-156-761-927	Sequence 927, App
37	4.6	20922	13	US-09-942-025-14	Sequence 14, Appl
38	4.6	23238	15	US-10-156-761-415	Sequence 415, App
39	4.6	30690	17	US-10-204-862A-1	Sequence 1, Appli
40	4.6	31422	17	US-10-204-862A-2	Sequence 2, Appli
41	4.6	67111	13	US-09-942-025-1	Sequence 1, Appli
42	4.6	80161	17	US-10-329-148A-1	Sequence 1, Appli
43	4.6	85692	17	US-10-461-194-1	Sequence 1, Appli
44	4.6	27705	13	US-09-942-025-12	Sequence 12, Appl
45	4.2	403	15	US-10-259-678-391	Sequence 391, App

ALIGNMENTS

RESULT 1

US-10-229-148B-1

Sequence 1, Application US/10229148B

Publication No. US20040091975A1

GENERAL INFORMATION:

APPLICANT: Meiji Seika Kaisha, Ltd.

TITLE OF INVENTION: Midcamycin biosynthetic genes

FILE REFERENCE: 138451 US

CURRENT APPLICATION NUMBER: US/10/229,148B

CURRENT FILING DATE: 2002-08-28

PRIOR APPLICATION NUMBER: 210516/2002

PRIOR FILING DATE: 2002-07-19

NUMBER OF SEQ ID NOS: 48

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 84428

TYPE: DNA

ORGANISM: Streptomyces mycarofaciens

FEATURE:

NAME/KEY: CDS

LOCATION: Complement((1)..(675))

OTHER INFORMATION: ORF42 (fragment)

FEATURE:

NAME/KEY: CDS

LOCATION: Complement((1168)..(2202))

OTHER INFORMATION: ORF41

FEATURE:

NAME/KEY: CDS

LOCATION: Complement((2220)..(3215))

OTHER INFORMATION: ORF40

FEATURE:

NAME/KEY: CDS

LOCATION: Complement((3237)..(4691))

OTHER INFORMATION: ORF39

FEATURE:

NAME/KEY: CDS

LOCATION: Complement((4695)..(5948))

```
OTHER INFORMATION: ORF38
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((6048)..(6629))
OTHER INFORMATION: ORF37
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((6653)..(7945))
OTHER INFORMATION: ORF36
FEATURE:
NAME/KEY: CDS
LOCATION: (8149)..(9015)
OTHER INFORMATION: ORF35
FEATURE:
NAME/KEY: CDS
LOCATION: (9012)..(9335)
OTHER INFORMATION: ORF34
FEATURE:
NAME/KEY: CDS
LOCATION: (9328)..(10458)
OTHER INFORMATION: ORF33
FEATURE:
NAME/KEY: CDS
LOCATION: (10521)..(11603)
OTHER INFORMATION: ORF32
FEATURE:
NAME/KEY: CDS
LOCATION: (11729)..(12961)
OTHER INFORMATION: ORF31
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((13016)..(14044))
OTHER INFORMATION: ORF30
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((14074)..(15096))
OTHER INFORMATION: ORF29
FEATURE:
NAME/KEY: CDS
LOCATION: (15643)..(17466)
OTHER INFORMATION: ORF28
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((17522)..(18895))
OTHER INFORMATION: ORF27
FEATURE:
NAME/KEY: CDS
LOCATION: (19063)..(20229)
OTHER INFORMATION: ORF26
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((20307)..(21743))
OTHER INFORMATION: ORF25
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((21733)..(22527))
OTHER INFORMATION: ORF24
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((22534)..(23571))
OTHER INFORMATION: ORF23
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((23555)..(24463))
OTHER INFORMATION: ORF22
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((24460)..(25650))
OTHER INFORMATION: ORF21
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((25647)..(26105))
OTHER INFORMATION: ORF20
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((26180)..(27391))
OTHER INFORMATION: ORF19
FEATURE:
NAME/KEY: CDS
LOCATION: (27937)..(28983)
OTHER INFORMATION: ORF18
FEATURE:
NAME/KEY: CDS
LOCATION: (29244)..(42779)
OTHER INFORMATION: ORF1
FEATURE:
NAME/KEY: CDS
LOCATION: (42823)..(48657)
OTHER INFORMATION: ORF2
FEATURE:
NAME/KEY: CDS
LOCATION: (48712)..(59802)
OTHER INFORMATION: ORF3
FEATURE:
NAME/KEY: CDS
LOCATION: (59850)..(64556)
OTHER INFORMATION: ORF4
FEATURE:
NAME/KEY: CDS
LOCATION: (64687)..(70365)
OTHER INFORMATION: ORF5
FEATURE:
NAME/KEY: CDS
LOCATION: (70365)..(71078)
OTHER INFORMATION: ORF6
FEATURE:
NAME/KEY: CDS
LOCATION: (71113)..(72360)
OTHER INFORMATION: ORF7
FEATURE:
NAME/KEY: CDS
LOCATION: (72400)..(73665)
OTHER INFORMATION: ORF8
FEATURE:
NAME/KEY: CDS
LOCATION: (73694)..(75043)
OTHER INFORMATION: ORF9
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((75899)..(76570))
OTHER INFORMATION: ORF10
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((76602)..(77765))
OTHER INFORMATION: ORF11
FEATURE:
NAME/KEY: CDS
LOCATION: (78039)..(79313)
OTHER INFORMATION: ORF12
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((79391)..(81052))
OTHER INFORMATION: ORF13
FEATURE:
NAME/KEY: CDS
LOCATION: (81541)..(82356)
OTHER INFORMATION: ORF14
FEATURE:
NAME/KEY: CDS
LOCATION: (82760)..(83362)
OTHER INFORMATION: ORF15
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((83495)..(84142))
OTHER INFORMATION: ORF16
FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (84329)..(84428)
; OTHER INFORMATION: ORF17 (fragment)
US-10-229-148B-1

Query Match          6.4%; Score 32; DB 17; Length 84428;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 136 GCCCCGACGCGCGTCAAGCGGTGCGCGA 167
Db 60600 GCCCCGACGCGCGTCAAGCGGTGCGCGA 60631

RESULT 2
US-10-156-761-412
; Sequence 412, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 412
; LENGTH: 18435
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(18435)
US-10-156-761-412

Query Match          5.8%; Score 29; DB 15; Length 18435;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 GACGTCGACGCGGTGAGCGCGCACGCGAC 410
Db 10033 GACGTCGACGCGGTGAGCGCGCACGCGAC 10061

RESULT 3
US-09-940-316B-1/c
; Sequence 1, Application US/09940316B
; Publication No. US20030175901A1
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, INC.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTO, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYPEPTIDES ENCODING THE FKBP GENE OF THE FK-520 POLYKETIDE SYNTH
; FILE REFERENCE: 30062-20026.11
; CURRENT APPLICATION NUMBER: US/09/940,316B
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/410,551
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
```

```
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275)...(71465)
US-09-940-316B-1

Query Match          5.8%; Score 29; DB 10; Length 77536;
Best Local Similarity 100.0%; Pred. No. 5.4e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 GCGGTCAACTCCGACGCGCGTCAAGCGG 305
Db 37516 GCGGTCAACTCCGACGCGCGTCAAGCGG 37488

RESULT 4
US-10-156-761-15103
; Sequence 15103, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15103
; LENGTH: 100000
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15103

Query Match          5.8%; Score 29; DB 15; Length 100000;
Best Local Similarity 100.0%; Pred. No. 5.1e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 GACGTCGACGCGGTGAGCGCGCACGCGAC 410
Db 60556 GACGTCGACGCGGTGAGCGCGCACGCGAC 60584

RESULT 5
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
```

; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-23  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 5.8%; Score 29; DB 15; Length 9025608;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 GAGTCGACGGGTGGAGGCGCAGGCAC 410  
|||  
Db 517356 GAGTCGACGGGTGGAGGCGCAGGCAC 517328

## RESULT 6

US-10-156-761-2885  
; Sequence 2885, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2885  
; LENGTH: 14520  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(14520)

US-10-156-761-2885

Query Match 5.4%; Score 27; DB 15; Length 14520;  
Best Local Similarity 100.0%; Pred. No. 0.00078;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 CCGACGGCGGTGCAAGGCGTTCGCGG 166  
|||  
Db 5042 CCGACGGCGGTGCAAGGCGTTCGCGG 5068

## RESULT 7

US-09-861-289-1  
; Sequence 1, Application US/09861289  
; Patent No. US20020110897A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/861,289  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-861-289-1

Query Match 5.4%; Score 27; DB 9; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.00076;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGGCGGCGG 131  
|||  
Db 3721 GTTCGTCGAGTTCTCGCGGCGGCGG 3747

## RESULT 8

US-09-860-846-1  
; Sequence 1, Application US/09860846  
; Patent No. US20020164742A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/860,846  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-860-846-1

Query Match 5.4%; Score 27; DB 9; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.00076;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGGCGGCGG 131  
|||  
Db 3721 GTTCGTCGAGTTCTCGCGGCGGCGG 3747

## RESULT 9

US-09-988-384B-1  
; Sequence 1, Application US/09988384B  
; Publication No. US20030073824A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.536US1  
; CURRENT APPLICATION NUMBER: US/09/988,384B  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: PCT/US99/14398  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 09/105,537

```
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

Query Match      5.4%; Score 27; DB 10; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 131
Db 3721 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 3747

RESULT 10
US-09-836-821-1
; Sequence 44, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-1

Query Match      5.4%; Score 27; DB 10; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 131
Db 3721 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 3747

RESULT 11
US-10-271-889-44
; Sequence 44, Application US/10271889
; Publication No. US20030194784A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin
; FILE REFERENCE: 600.582US1
; CURRENT APPLICATION NUMBER: US/10/271,889
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44

; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-10-271-889-44

Query Match      5.4%; Score 27; DB 15; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 131
Db 3721 GTTCGTCGAGTTCTCGCGGCGAGCGCGG 3747

RESULT 12
US-10-156-761-15102
; Sequence 15102, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15102
; LENGTH: 125746
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15102

Query Match      5.4%; Score 27; DB 15; Length 125746;
Best Local Similarity 100.0%; Pred. No. 0.00048;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 140 CCGACGGCGGCTGCAAGCGGTTTCGCGG 166
Db 48818 CCGACGGCGGCTGCAAGCGGTTTCGCGG 48844

RESULT 13
US-10-156-761-414
; Sequence 414, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 414
; LENGTH: 10692
```



```
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(10692)
US-10-156-761-414

Query Match          5.2%; Score 26; DB 15; Length 10692;
Best Local Similarity 100.0%; Pred.No. 0.0026;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 GTCGACGCGTGGAGGCCACAGGCAC 410
Db 1003 GTCGACGCGTGGAGGCCACAGGCAC 1028

RESULT 14
US-10-378-083-1
; Sequence 1, Application US/10378083
; Publication No. US20040053274A1
; GENERAL INFORMATION:
; APPLICANT: President of Tokyo Institute of Technology
; TITLE OF INVENTION: A gene cluster of vicenistatin biosynthase, a vicenistamine
; TITLE OF INVENTION: glycosyltransferase polypeptide, and a gene encoding the
; TITLE OF INVENTION: polypeptide
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/378,083
; CURRENT FILING DATE: 2003-03-04
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 64492
; TYPE: DNA
; ORGANISM: Streptomyces halstedii sp.HC-34
US-10-378-083-1

Query Match          5.2%; Score 26; DB 13; Length 64492;
Best Local Similarity 100.0%; Pred.No. 0.0017;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 AGGCGCTGCTGGCCACCTACGGCCAG 468
Db 54630 AGGCGCTGCTGGCCACCTACGGCCAG 54655

RESULT 15
US-09-808-880-1
; Sequence 1, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-808-880-1
```

```
Query Match          5.0%; Score 25; DB 10; Length 50937;
Best Local Similarity 100.0%; Pred.No. 0.0058;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 GACGGCGCTGCAAGGCGTTCGCCG 166
Db 9582 GACGGCGCTGCAAGGCGTTCGCCG 9606

RESULT 16
US-09-953-348-19
; Sequence 19, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingqing
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953,348
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: PCT/US00/06394
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/266965
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 3978
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-953-348-19
```

```
Query Match          4.8%; Score 24; DB 10; Length 3978;
Best Local Similarity 100.0%; Pred.No. 0.032;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 CTGGCCCCCGACGGCGCTGCAAG 156
Db 757 CTGGCCCCCGACGGCGCTGCAAG 780

RESULT 17
US-10-267-255-19
; Sequence 19, Application US/10267255
; Publication No. US20030124689A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 3978
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
```

```

US-10-267-255-19
Query Match          4.8%; Score 24; DB 15; Length 3978;
Best Local Similarity 100.0%; Pred. No. 0.032;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCGAGCGCGCTGCAAG 156
      |||||
DB 757 CTGGCCCCGAGCGCGCTGCAAG 780

RESULT 18
US-10-156-761-413
; Sequence 413, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 413
; LENGTH: 5505
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5505)
US-10-156-761-413

Query Match          4.8%; Score 24; DB 15; Length 5505;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 CGCGAGTGCTCCATGGCGCTGCG 71
      |||||
DB 663 CGCGAGTGCTCCATGGCGCTGCG 686

RESULT 19
US-10-156-761-2880
; Sequence 2880, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2880
; LENGTH: 5721
US-10-156-761-2880

Query Match          4.8%; Score 24; DB 15; Length 11910;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 AGCGCTGCTGGCCACCTACGGCC 466
      |||||
DB 1022 AGCGCTGCTGGCCACCTACGGCC 1045

RESULT 21
US-09-953-348-74
; Sequence 74, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingqing
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953,348
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: PCT/US00/06394
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/266965
; PRIOR FILING DATE: 1999-03-12
US-09-953-348-74

Query Match          4.8%; Score 24; DB 15; Length 11910;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 AGCGCTGCTGGCCACCTACGGCC 466
      |||||
DB 1022 AGCGCTGCTGGCCACCTACGGCC 1045

RESULT 21
US-09-953-348-74
; Sequence 74, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingqing
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953,348
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: PCT/US00/06394
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/266965
; PRIOR FILING DATE: 1999-03-12
US-09-953-348-74

```

; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 74  
; LENGTH: 12249  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-953-348-74

Query Match 4.8%; Score 24; DB 10; Length 12249;  
Best Local Similarity 100.0%; Pred. No. 0.025;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCGACGGCGCGCTGCAAG 156  
|||  
Db 1744 CTGGCCCCGACGGCGCGCTGCAAG 1767

## RESULT 22

US-10-267-255-74

; Sequence 74, Application US/10267255  
; Publication No. US20030124689A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456US1  
; CURRENT APPLICATION NUMBER: US/10/267,255  
; CURRENT FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US 09/266,965  
; PRIOR FILING DATE: 1999-03-12  
; PRIOR APPLICATION NUMBER: US 08/624,447  
; PRIOR FILING DATE: 1996-08-19  
; PRIOR APPLICATION NUMBER: PCT/US94/11279  
; PRIOR FILING DATE: 1994-10-06  
; PRIOR APPLICATION NUMBER: US 08/133,963  
; PRIOR FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 74  
; LENGTH: 12249  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-10-267-255-74

Query Match 4.8%; Score 24; DB 15; Length 12249;  
Best Local Similarity 100.0%; Pred. No. 0.025;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCGACGGCGCGCTGCAAG 156  
|||  
Db 1744 CTGGCCCCGACGGCGCGCTGCAAG 1767

## RESULT 23

US-09-953-348-96

; Sequence 96, Application US/09953348  
; Publication No. US20030134398A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, David, H  
; APPLICANT: Mao, Yingqing  
; APPLICANT: Varoglu, Mustafa  
; APPLICANT: He, Min  
; APPLICANT: Sheldon, Paul  
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER  
; FILE REFERENCE: 600.530US1  
; CURRENT APPLICATION NUMBER: US/09/953,348  
; CURRENT FILING DATE: 2001-09-12  
; PRIOR APPLICATION NUMBER: PCT/US00/06394  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/266965

; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 96  
; LENGTH: 18331  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-953-348-96

Query Match 4.8%; Score 24; DB 10; Length 18331;  
Best Local Similarity 100.0%; Pred. No. 0.023;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCGACGGCGCGCTGCAAG 156  
|||  
Db 1744 CTGGCCCCGACGGCGCGCTGCAAG 1767

## RESULT 24

US-10-267-255-96

; Sequence 96, Application US/10267255  
; Publication No. US20030124689A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456US1  
; CURRENT APPLICATION NUMBER: US/10/267,255  
; CURRENT FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US 09/266,965  
; PRIOR FILING DATE: 1999-03-12  
; PRIOR APPLICATION NUMBER: US 08/624,447  
; PRIOR FILING DATE: 1996-08-19  
; PRIOR APPLICATION NUMBER: PCT/US94/11279  
; PRIOR FILING DATE: 1994-10-06  
; PRIOR APPLICATION NUMBER: US 08/133,963  
; PRIOR FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 96  
; LENGTH: 18331  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-10-267-255-96

Query Match 4.8%; Score 24; DB 15; Length 18331;  
Best Local Similarity 100.0%; Pred. No. 0.023;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCGACGGCGCGCTGCAAG 156  
|||  
Db 1744 CTGGCCCCGACGGCGCGCTGCAAG 1767

## RESULT 25

US-10-461-194-2

; Sequence 2, Application US/10461194  
; Publication No. US20040077058A1  
; GENERAL INFORMATION:  
; APPLICANT: Hutchinson, Richard C.  
; APPLICANT: Reid, Ralph C.  
; APPLICANT: Hu, Zhihao  
; APPLICANT: Rascher, Andreas  
; APPLICANT: Schirmer, Andreas  
; APPLICANT: McDaniel, Robert  
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
; TITLE OF INVENTION: PRO-CELLANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND  
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF  
; FILE REFERENCE: 300622009700  
; CURRENT APPLICATION NUMBER: US/10/461,194

; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389,255  
; PRIOR FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/393,929  
; PRIOR FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/395,275  
; PRIOR FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: US 10/212,962  
; PRIOR FILING DATE: 2002-08-05  
; PRIOR APPLICATION NUMBER: US 60/415,326  
; PRIOR FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: US 60/420,820  
; PRIOR FILING DATE: 2002-10-24  
; PRIOR APPLICATION NUMBER: US 60/433,130  
; PRIOR FILING DATE: 2002-12-13  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 86941  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus  
US-10-461-194-2

Query Match 4.8%; Score 24; DB 17; Length 86941;  
Best Local Similarity 100.0%; Pred. No. 0.016; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 0

QY 445 GCGCTGCTGGCCACCTACGGCCAG 468  
DB 54584 GCGCTGCTGGCCACCTACGGCCAG 54607  
|||||

RESULT 26  
US-10-156-761-15102/c  
; Sequence 15102, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 15102  
; LENGTH: 125746  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-15102

Query Match 4.8%; Score 24; DB 15; Length 125746;  
Best Local Similarity 100.0%; Pred. No. 0.015; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 0

QY 443 AGCGCTGCTGGCCACCTACGGCC 466  
DB 113871 AGCGCTGCTGGCCACCTACGGCC 113848  
|||||

RESULT 27  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 4.8%; Score 24; DB 15; Length 9025608;  
Best Local Similarity 100.0%; Pred. No. 0.0056; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 0

QY 443 AGCGCTGCTGGCCACCTACGGCC 466  
DB 3537788 AGCGCTGCTGGCCACCTACGGCC 3537811  
|||||

RESULT 28  
US-09-940-316B-26  
; Sequence 26, Application US/09940316B  
; Publication No. US20030175901A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDES ENCODING THE FKBA GENE OF THE FK-520 POLYKETIDE SYNTHASE  
; FILE REFERENCE: 30062-20026.11  
; CURRENT APPLICATION NUMBER: US/09/940,316B  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 09/410,551  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26  
; LENGTH: 4674  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; OTHER INFORMATION: PKS synthase fragment  
; NAME/KEY: CDS  
; LOCATION: (3)..(4673)  
US-09-940-316B-26

Query Match 4.6%; Score 23; DB 10; Length 4674;  
Best Local Similarity 100.0%; Pred. No. 0.097;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGGTGCAACGG 305  
|||||  
DB 1254 AACTCCGACGGCGGTGCAACGG 1276

## RESULT 29

US-09-940-316B-24  
; Sequence 24, Application US/09940316B  
; Publication No. US20030175901A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTU, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH  
; TITLE OF INVENTION: GENE CLUSTER  
; FILE REFERENCE: 30062-20026.11  
; CURRENT APPLICATION NUMBER: US/09/940,316B  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 09/410,551  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 24  
; LENGTH: 4725

TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic

OTHER INFORMATION: PKS synthase fragment  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3)..(4724)

US-09-940-316B-24

Query Match 4.6%; Score 23; DB 10; Length 4725;  
Best Local Similarity 100.0%; Pred. No. 0.097;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGGTGCAACGG 305  
|||||  
DB 1254 AACTCCGACGGCGGTGCAACGG 1276

## RESULT 30

US-09-940-316B-30  
; Sequence 30, Application US/09940316B  
; Publication No. US20030175901A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTU, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH  
; TITLE OF INVENTION: GENE CLUSTER  
; FILE REFERENCE: 30062-20026.11  
; CURRENT APPLICATION NUMBER: US/09/940,316B  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 09/410,551

;; PRIOR FILING DATE: 1999-10-01  
;; PRIOR APPLICATION NUMBER: US 60/139,650  
;; PRIOR FILING DATE: 1999-06-17  
;; PRIOR APPLICATION NUMBER: US 60/123,810  
;; PRIOR FILING DATE: 1999-03-11  
;; PRIOR APPLICATION NUMBER: US 60/102,748  
;; PRIOR FILING DATE: 1998-10-02  
;; NUMBER OF SEQ ID NOS: 72  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 30  
;; LENGTH: 4737  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
;; OTHER INFORMATION: PKS synthase fragment  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (3)..(4736)  
US-09-940-316B-30

Query Match 4.6%; Score 23; DB 10; Length 4737;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGGTGCAACGG 305  
|||||  
DB 1254 AACTCCGACGGCGGTGCAACGG 1276

## RESULT 31

US-09-940-316B-28  
; Sequence 28, Application US/09940316B  
; Publication No. US20030175901A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTU, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH  
; TITLE OF INVENTION: GENE CLUSTER  
; FILE REFERENCE: 30062-20026.11  
; CURRENT APPLICATION NUMBER: US/09/940,316B  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 09/410,551  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 28  
; LENGTH: 4767  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; OTHER INFORMATION: PKS synthase fragment  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)..(4766)  
US-09-940-316B-28

Query Match 4.6%; Score 23; DB 10; Length 4767;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGGTGCAACGG 305

Db 1254 AACTCCGACGGCGTCAACGG 1276  
|||||

## RESULT 32

US-09-940-316B-32  
; Sequence 32, Application US/09940316B  
; Publication No. US20030175901A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDES ENCODING THE FKbA GENE OF THE FK-520 POLYKETIDE SYNTHASE  
; FILE REFERENCE: 30062-20026.11  
; CURRENT APPLICATION NUMBER: US/09/940,316B  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 09/410,551  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 4818  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; OTHER INFORMATION: PKS synthase fragment  
; NAME/KEY: CDS  
; LOCATION: (3)..(4817)  
US-09-940-316B-32

Query Match 4.6%; Score 23; DB 10; Length 4818;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 RACTCCGACGGCGTCAACGG 305  
Db 1254 RACTCCGACGGCGTCAACGG 1276  
|||||

## RESULT 33

US-10-156-761-411  
; Sequence 411, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 411

; LENGTH: 10056  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(10056)  
US-10-156-761-411

Query Match 4.6%; Score 23; DB 15; Length 10056;  
Best Local Similarity 100.0%; Pred. No. 0.081;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 CCGACGGCGCTGCAAGGGTTC 162  
Db 5237 CCGACGGCGCTGCAAGGGTTC 5259  
|||||

## RESULT 34

US-10-156-761-2883  
; Sequence 2883, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2883  
; LENGTH: 14055  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(14055)  
US-10-156-761-2883

Query Match 4.6%; Score 23; DB 15; Length 14055;  
Best Local Similarity 100.0%; Pred. No. 0.076;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 CTGGGCGACCCGATCGAGGCCGA 443  
Db 1012 CTGGGCGACCCGATCGAGGCCGA 1034  
|||||

## RESULT 35

US-10-156-761-931  
; Sequence 931, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 931  
; LENGTH: 14643  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; NAME/KEY: CDS  
; LOCATION: (1)..(14643)  
US-10-156-761-931

Query Match 4.6%; Score 23; DB 15; Length 14643;  
Best Local Similarity 100.0%; Pred. No. 0.075;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 TGGTGGGCGGCTGTCCGAGGCT 234  
|||||  
DB 5501 TGGTGGGCGGCTGTCCGAGGCT 5523  
|||||

## RESULT 36

US-10-156-761-927  
; Sequence 927, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 927  
; LENGTH: 18717  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(18717)  
US-10-156-761-927

Query Match 4.6%; Score 23; DB 15; Length 18717;  
Best Local Similarity 100.0%; Pred. No. 0.071;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 418 ACCCTGGGCGACCCGATCGAGGC 440  
|||||  
DB 1045 ACCCTGGGCGACCCGATCGAGGC 1067  
|||||

## RESULT 37

US-09-942-025-14  
; Sequence 14, Application US/09942025  
; Publication No. US20030054547A1  
; GENERAL INFORMATION:  
; APPLICANT: Kosan Biosciences, Inc.  
; APPLICANT: Julien, Bryan  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENE FROM SORANGIUM  
; FILE REFERENCE: 30062-20020.21  
; CURRENT APPLICATION NUMBER: US/09/942,025  
; CURRENT FILING DATE: 2001-08-28

; PRIOR APPLICATION NUMBER: US 60/271,245  
; PRIOR FILING DATE: 2001-02-15  
; PRIOR APPLICATION NUMBER: US 09/144,085  
; PRIOR FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: US 09/010,809  
; PRIOR FILING DATE: 1998-01-22  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 20922  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-942-025-14

Query Match 4.6%; Score 23; DB 13; Length 20922;  
Best Local Similarity 100.0%; Pred. No. 0.069;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 420 CCTGGGCGACCCGATCGAGGCG 442  
|||||  
DB 1032 CCTGGGCGACCCGATCGAGGCG 1054  
|||||

## RESULT 38

US-10-156-761-415  
; Sequence 415, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 415  
; LENGTH: 23238  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(23238)  
US-10-156-761-415

Query Match 4.6%; Score 23; DB 15; Length 23238;  
Best Local Similarity 100.0%; Pred. No. 0.068;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 GACGGCGCTGCAAGCGTTGCG 164  
|||||  
DB 3796 GACGGCGCTGCAAGCGTTGCG 3818  
|||||

## RESULT 39

US-10-204-862A-1  
; Sequence 1, Application US/10204862A  
; Publication No. US20040101936A1  
; GENERAL INFORMATION:  
; APPLICANT: ENDO, HIROFUMI  
; APPLICANT: YAMAGUCHI, HIROYUKI  
; APPLICANT: KANDA, YUTAKA  
; APPLICANT: HASHIMOTO, SHINICHI  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO



## ; TITLE OF INVENTION: PROCESS FOR PRODUCING AVERMECTIN DERIVATIVE

; FILE REFERENCE: 468-32

; CURRENT APPLICATION NUMBER: US/10/204,862A

; CURRENT FILING DATE: 2003-12-26

; PRIOR APPLICATION NUMBER: JP 00/047405

; PRIOR FILING DATE: 2000-02-24

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 30690

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(11916)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (11971)..(30687)

; US-10-204-862A-1

Query Match 4.6%; Score 23; DB 17; Length 30690;

Best Local Similarity 100.0%; Pred. No. 0.063; 0; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 418 ACCCTGGGCGACCCGATCGAGGC 440

Db 13015 ACCCTGGGCGACCCGATCGAGGC 13037

## RESULT 40

US-10-204-862A-2

; Sequence 2, Application US/10204862A

; Publication No. US20040101936A1

; GENERAL INFORMATION:

; APPLICANT: ENDO, HIROFUMI

; APPLICANT: YAMAGUCHI, HIROYUKI

; APPLICANT: KANDA, YUTAKA

; APPLICANT: HASHIMOTO, SHINICHI

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; TITLE OF INVENTION: PROCESS FOR PRODUCING AVERMECTIN DERIVATIVE

; FILE REFERENCE: 468-32

; CURRENT APPLICATION NUMBER: US/10/204,862A

; CURRENT FILING DATE: 2003-12-26

; PRIOR APPLICATION NUMBER: JP 00/047405

; PRIOR FILING DATE: 2000-02-24

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 31422

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(14643)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (14824)..(31419)

; US-10-204-862A-2

Query Match 4.6%; Score 23; DB 17; Length 31422;

Best Local Similarity 100.0%; Pred. No. 0.063; 0; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 TGGTGGAGCGGCTGTCCGAGGCT 234

Db 5501 TGGTGGAGCGGCTGTCCGAGGCT 5523

## RESULT 41

US-09-942-025-1

; Sequence 1, Application US/09942025

; Publication No. US20030054547A1

## ; GENERAL INFORMATION:

; APPLICANT: Kosan Biosciences, Inc.

; APPLICANT: Julien, Bryan

; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENE FROM SORANGIUM

; TITLE OF INVENTION: CELLULOSUM

; FILE REFERENCE: 30062-20020.21

; CURRENT APPLICATION NUMBER: US/09/942,025

; CURRENT FILING DATE: 2001-08-28

; PRIOR APPLICATION NUMBER: US 60/271,245

; PRIOR FILING DATE: 2001-02-15

; PRIOR APPLICATION NUMBER: US 09/144,085

; PRIOR FILING DATE: 1998-08-31

; PRIOR APPLICATION NUMBER: US 09/010,809

; PRIOR FILING DATE: 1998-01-22

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 67311

; TYPE: DNA

; ORGANISM: Sorangium cellulosum

; US-09-942-025-1

Query Match 4.6%; Score 23; DB 13; Length 67311;

Best Local Similarity 100.0%; Pred. No. 0.053; 0; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 420 CCTGGGCGACCCGATCGAGGCG 442

Db 32581 CCTGGGCGACCCGATCGAGGCG 32603

## RESULT 42

US-10-329-148A-1

; Sequence 1, Application US/10329148A

; Publication No. US20040023343A1

; GENERAL INFORMATION:

; APPLICANT: Baltz, Richard H

; APPLICANT: Broughton, Mary C

; APPLICANT: Crawford, Kathryn P

; APPLICANT: Madduri, Krishnamurthy

; APPLICANT: Treadway, Patti J

; APPLICANT: Turner, Jan R

; APPLICANT: Waldron, Clive

; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide

; FILE REFERENCE: 50489 DIV1

; CURRENT APPLICATION NUMBER: US/10/329,148A

; CURRENT FILING DATE: 2002-12-23

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/603,207B

; PRIOR FILING DATE: EARLIER FILING DATE: 2000-06-23

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/370,700

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-09

; NUMBER OF SEQ ID NOS: 39

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 80161

; TYPE: DNA

; ORGANISM: Saccharopolyspora spinosa

; US-10-329-148A-1

Query Match 4.6%; Score 23; DB 17; Length 80161;

Best Local Similarity 100.0%; Pred. No. 0.051; 0; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GACGGCACCGGCTGGTCCGAGGG 197

Db 24771 GACGGCACCGGCTGGTCCGAGGG 24793

## RESULT 43

US-10-461-194-1

; Sequence 1, Application US/10461194

; Publication No. US20040077058A1

; GENERAL INFORMATION:

```

; APPLICANT: Hutchinson, Richard C.
; APPLICANT: Reid, Ralph C.
; APPLICANT: Hu, Zhihao
; APPLICANT: Rascher, Andreas
; APPLICANT: Schirmer, Andreas
; APPLICANT: McDaniel, Robert
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
; TITLE OF INVENTION: PRO-CELLULOSE PRODUCING POLYKETIDE SYNTHASES AND
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF
; FILE REFERENCE: 300622009700
; CURRENT APPLICATION NUMBER: US/10/461,194
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,255
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/393,929
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/395,275
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/415,326
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/420,820
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/433,130
; PRIOR FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 85692
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-461-194-1

```

```

Query Match          4.6%; Score 23; DB 17; Length 85692;
Best Local Similarity 100.0%; Pred. No. 0.05;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      451 CTGGCCACTAGCGGCGAGACCG 473
DB      49924 CTGGCCACTAGCGGCGAGACCG 49946

```

```

RESULT 44
US-09-942-025-12
; Sequence 12, Application US/09942025
; Publication No. US20030054547A1
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENE FROM SORANGIUM
; TITLE OF INVENTION: CELLULOSE
; FILE REFERENCE: 30062-20020.21
; CURRENT APPLICATION NUMBER: US/09/942,025
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/271,245
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: US 09/144,085
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: US 09/010,809
; PRIOR FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 27705
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-942-025-12

```

```

Query Match          4.4%; Score 22; DB 13; Length 27705;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      386 TCGACGGCGGTGGAGGCGCACGG 407
DB      13082 TCGACGGCGGTGGAGGCGCACGG 13103

```

```

RESULT 45
US-10-259-678-391/c
; Sequence 391, Application US/10259678
; Publication No. US20030198974A1
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/10/259,678
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 391
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-10-259-678-391

```

```

Query Match          4.2%; Score 21; DB 15; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      393 GGTGGAGGCGGCGCACGG 413
DB      383 GGTGGAGGCGGCGCACGG 363

```

```

Search completed: June 16, 2004, 20:21:25
Job time : 1342 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:37:20 ; Search time 2102.33 seconds  
(without alignments)  
7116.356 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934

Perfect score: 501

Sequence: 1 tcgtcttcgctggcgcgcgt.....cgctgtggtcggtcggtg 501

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 15

Total number of hits satisfying chosen parameters: 15169

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estmu.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_hic.\*
- 9: gb\_est1.\*
- 10: gb\_est2.\*
- 11: gb\_hic.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: em\_estfun.\*
- 16: em\_estcom.\*
- 17: em\_gss\_hum.\*
- 18: em\_gss\_inv.\*
- 19: em\_gss\_pln.\*
- 20: em\_gss\_vrt.\*
- 21: em\_gss\_fun.\*
- 22: em\_gss\_man.\*
- 23: em\_gss\_mus.\*
- 24: em\_gss\_pro.\*
- 25: em\_gss\_rod.\*
- 26: em\_gss\_pig.\*
- 27: em\_gss\_vrl.\*
- 28: gb\_gss1.\*
- 29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	4.4	361	13	BY090167
2	22	4.4	574	12	BM397194
3	21	4.2	159	9	AA014791
4	21	4.2	247	14	W65007

5	21	4.2	253	9	AA509737	AA509737
6	21	4.2	264	13	BY074439	BY074439
7	21	4.2	309	14	W41546	W41546
8	21	4.2	322	14	W41172	W41172
9	21	4.2	325	9	AA139922	AA139922
10	21	4.2	333	13	BY055063	BY055063
11	21	4.2	335	13	BY052633	BY052633
12	21	4.2	337	9	AA052685	AA052685
13	21	4.2	338	10	BB869636	BB869636
14	21	4.2	345	10	BE859603	BE859603
15	21	4.2	366	13	BY168519	BY168519
16	21	4.2	371	13	BY078228	BY078228
17	21	4.2	372	10	BB871048	BB871048
18	21	4.2	384	14	W48899	W48899
19	21	4.2	396	9	AA914086	AA914086
20	21	4.2	419	9	AA051350	AA051350
21	21	4.2	421	13	BY227782	BY227782
22	21	4.2	448	9	AA020238	AA020238
23	21	4.2	489	14	CA600107	CA600107
24	21	4.2	501	9	AA790556	AA790556
25	21	4.2	544	10	BF581564	BF581564
26	21	4.2	548	9	AA711986	AA711986
27	21	4.2	565	12	BM070219	BM070219
28	21	4.2	573	13	BU582780	BU582780
29	21	4.2	591	9	AA265232	AA265232
30	21	4.2	627	29	CE288123	CE288123
31	21	4.2	644	10	BE286789	BE286789
32	21	4.2	670	13	BY703379	BY703379
33	21	4.2	673	13	BU757970	BU757970
34	21	4.2	686	11	AK002881	AK002881
35	21	4.2	698	13	BY737644	BY737644
36	21	4.2	698	10	BE371234	BE371234
37	21	4.2	702	13	BY709009	BY709009
38	21	4.2	714	13	BY710477	BY710477
39	21	4.2	715	11	AK011122	AK011122
40	21	4.2	728	11	AK009191	AK009191
41	21	4.2	733	12	BI687681	BI687681
42	21	4.2	747	10	BE282094	BE282094
43	21	4.2	761	12	BG976291	BG976291
44	21	4.2	826	13	BU962156	BU962156
45	21	4.2	854	14	CB574912	CB574912

ALIGNMENTS

RESULT 1

BY090167

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BY090167 361 bp mRNA linear EST 07-DEC-2002  
BY090167 RIKEN full-length enriched, pooled tissues, adult spleen,  
etc. Mus musculus cDNA clone K630080E08 5', mRNA sequence.

BY090167 GI:26199455

EST.

Mus musculus

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 361)

Okazaki, F., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,

Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,

Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,

Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,

Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Bruscia, V.,

Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragana, I. A.,

Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,

Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,

Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,

Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,

Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,

Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G.,

Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reid,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Sempile,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L.G., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

Analysis of the mouse transcriptome based on functional annotation

Of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,

Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,

Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,

Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,

Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and

Hayashizaki,Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in Riken.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for

further details.

Location/Qualifiers

1. .361

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="X630080E08"

/clone\_lib="RIKEN full-length enriched, pooled tissues,

adult spleen, etc."

/note="(dev stage=adult,tissue type=spleen,sex=male),

(dev stage=adult,tissue type=kidney,sex=male),

(dev stage=adult,tissue type=testis,sex=male),

(dev stage=adult,tissue type=thymus,sex=male),

(dev stage=adult,tissue type=heart,sex=male),

(dev stage=adult,tissue type=colon,sex=male),

(dev stage=adult,tissue type=stomach,sex=male),

(dev stage=adult,tissue type=liver,sex=male),

(dev stage=13 days embryo,tissue type=whole

body,sex=mix), (dev stage=14 days embryo,tissue type=whole

body,sex=mix), (dev stage=16 days embryo,tissue type=whole

body,sex=mix), (dev stage=17 days embryo,tissue type=whole

body,sex=mix), (dev stage=15 days pregnant,

adult,tissue type=amnion,sex=female), (dev stage=10 days

# ORIGIN

Query Match 4.4%; Score 22; DB 13; Length 361;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 210 CCTGCTGGAGCGGCTGTCCGAG 231  
DB 64 CCTGCTGGAGCGGCTGTCCGAG 85

## RESULT 2

BM397194/c  
LOCUS BM397194.1 574 bp mRNA linear EST 17-JAN-2002  
DEFINITION 5099-0-3-B07.t.1 Chilcoat/turkewitz cDNA (large fraction)  
Tetrahymena thermophila cDNA, mRNA sequence.  
ACCESSION BM397194.1 GI:18197247  
VERSION EST.  
KEYWORDS Tetrahymena thermophila  
SOURCE Tetrahymena thermophila  
ORGANISM Tetrahymena thermophila  
REFERENCE 1 (bases 1 to 574)  
AUTHORS Turkewitz,A.P., Karrer,K.M., Jahn,K., Orlas,E., Kirk,K.E.,  
Frankel,J., and Klobutcher,L.  
TITLE EST from Tetrahymena thermophila, strain CU428.1, growing cells  
JOURNAL Unpublished (2002)  
COMMENT Molecular Genetics and Cell Biology  
University of Chicago  
920 E. 58th Street, Chicago, IL 60637, USA  
Tel: 773 702 4374  
Fax: 773 702 3172  
Email: apturkew@midway.uchicago.edu  
Seq primer: T3.

## FEATURES

source  
1. .574  
/organism="Tetrahymena thermophila"  
/mol\_type="mRNA"  
/strain="CU428.1"  
/db\_xref="taxon:5911"  
/clone\_lib="Chilcoat/Turkewitz cDNA (large fraction)"  
/note="Vector: BlueScript 2 SK+; Details on library  
preparation can be found in Chilcoat and Turkewitz (2001)  
Proc. Natl. Acad. Sci USA, 98: 8709-8713."

## ORIGIN

Query Match 4.4%; Score 22; DB 12; Length 574;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 61 ATGGCGCTCGCGGCGCGGTGA 82  
DB 196 ATGGCGCTCGCGGCGCGGTGA 175

## RESULT 3

AA014791  
LOCUS AA014791 159 bp mRNA linear EST 21-JAN-1997  
DEFINITION m18G02.r1 Soares mouse placenta 4NBMP13.5 14.5 Mus musculus cDNA  
clone IMAGE:442898 5', mRNA sequence.  
ACCESSION AA014791  
VERSION AA014791.1 GI:1475654  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 159)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LINL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:268234  
Seq primer: mob.REGA+ET  
High quality sequence stop: 155.  
Location/Qualifiers  
1. .159  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:442898"  
/sex="unknown"  
/tissue\_type="placenta"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse placenta 4NbMPl3.5 14.5"  
/note="Organ: placenta; vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5, 5']  
TGTTCACATCTGAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT  
T 3'; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN  
Query Match 4.2%; Score 21; DB 9; Length 159;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231  
|||||  
Db 48 CTGGTGGAGCGGCTGTCCGAG 68  
|||||

RESULT 4  
W65007 247 bp mRNA linear EST 10-JUN-1996  
LOCUS me04a11.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA  
DEFINITION clone IMAGE:386492 5', mRNA sequence.  
ACCESSION W65007  
VERSION W65007.1 GI:1372649  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 247)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LINL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:248324  
Putative full length read  
Seq primer: ETPRimer  
High quality sequence stop: 238.  
Location/Qualifiers  
1. .247  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:386492"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse embryo NbME13.5 14.5"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5, 5']  
TGTTCACATCTGAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT  
T 3'; on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN  
Query Match 4.2%; Score 21; DB 14; Length 247;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231  
|||||  
Db 75 CTGGTGGAGCGGCTGTCCGAG 95  
|||||

RESULT 5  
AA509737 253 bp mRNA linear EST 08-JUL-1997  
LOCUS vH52a09.r1 Soares mammary\_gland NbVMG Mus musculus cDNA clone  
DEFINITION IMAGE:890584 5', mRNA sequence.  
ACCESSION AA509737  
VERSION AA509737.1 GI:2247591  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 253)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:518544

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 239.

Location/Qualifiers

#### FEATURES

source

1. 253  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:890584"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stages="4 weeks"  
/lab\_host="DH10B"  
/clone\_lib="Soares mammary gland NBMWG"  
/note="Organ: mammary gland; Vector: pT7T3D-Pac  
(Pharmacia) with a modified polylinker; Site 1: Not I;  
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo (dT) primer [5'  
TGTACCAATCTGAAGTCGAGCGGCCGGAATGTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M. Fatima  
Bonaldo."

#### ORIGIN

Query Match 4.2%; Score 21; DB 9; Length 253;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 CTGGTGGAGCGGCTGTCCGAG 231

Db 61 CTGGTGGAGCGGCTGTCCGAG 81

#### RESULT 6

BY074439

LOCUS

DEFINITION  
musculus cDNA clone 1920094L17 5', mRNA sequence.

ACCESSION

BY074439.1 GI:26176234

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 264)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Nikaido, I., Oatso, N., Saito, R., Suzuki, H., Yamanaka, I.,

Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,

Schönbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,

Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,

Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,

Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,

Fletcher, C. P., Forrest, A., Frazer, K. S., Gaasterland, T.,

Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,

Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,

Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. F., Konagaya, A.,

Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,

Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,

Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G.,

Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,

Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,

Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,

Sultana, C., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,

Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y.,

Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,

Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
Sakayama, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

#### TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, N., Waki, K., Watahiki, A., Muramatsu, M. and  
Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

#### FEATURES

source

1. 264

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CS7BL/6J"

/db\_xref="taxon:10090"

/clone="I920094L17"

/tissue\_type="kidney"

/dev\_stages="17 days embryo"

/clone\_lib="RIKEN full-length enriched, 17 days embryo  
kidney"

#### ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 264;

Best Local Similarity 100.0%; Pred. No. 3.3e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 CTGGTGGAGCGGCTGTCCGAG 231

Db 109 CTGGTGGAGCGGCTGTCCGAG 129

#### RESULT 7

W41546

LOCUS

DEFINITION

IMAGE:351932 5', mRNA sequence.

ACCESSION

W41546

W41546 309 bp mRNA linear EST 11-SEP-1996

mc50a11.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone





Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:355601  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 315.  
 Location/Qualifiers  
 1..325  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="BALB/c"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:580953"  
 /sex="mixed"  
 /tissue type="kidney"  
 /dev stage="6 weeks"  
 /lab\_host="DH10B"  
 /clone\_lib="Barstead MPLRB1"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5'  
 TGTTGCAATCTGAAGTGGAGCGGCGCTTTTTTTTTTTTTTTT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 [CATGATTCGTACC], digested with Not I and cloned into the  
 Not I and Eco RI sites of the modified pT73 vector.  
 Library constructed by Bob Barstead."

ORIGIN  
 Query Match 4.2%; Score 21; DB 9; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231  
 |||||  
 Db 83 CTGGTGGAGCGGCTGTCCGAG 103  
 |||||

RESULT 10  
 BY055063  
 LOCUS  
 DEFINITION  
 BY055063 333 bp mRNA linear EST 06-DEC-2002  
 CDNA clone 1730080N15 5', mRNA sequence.  
 BY055063  
 VERSION  
 KEYWORDS  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 333)  
 Okazaki.Y., Furuno.M., Kasukawa.T., Adachi.J., Bono.H., Kondo.S.,  
 Nikaido.I., Osato.N., Saito.R., Suzuki.H., Yamana.K.I.,  
 Kiyosawa.H., Yagi.K., Tomaru.Y., Hasegawa.Y., Nogami.A.,  
 Schonbach.C., Gojobori.T., Balderelli.R., Hill.D.P., Bult.C.,  
 Hume.D.A., Quackenbush.J., Schriml.L.M., Kanapin.A., Matsuda.H.,  
 Batalov.S., Beisel.K.W., Blake.J.A., Bradt.D., Brusic.V.,  
 Chothia.C., Corbani.L.E., Cousins.S., Dalla.E., Dragani.T.A.,  
 Fletcher.C.P., Forrest.A., Frazer.K.S., Gaasterland.T.,  
 Gariboldi.M., Gissi.C., Godzik.A., Gough.J., Grimmond.S.,  
 Gustincich.S., Hirokawa.N., Jackson.I.J., Jarvis.E.D., Kanai.A.,  
 Kawaji.H., Kawasawa.Y., Kedierski.R.M., King.B.H., Konagaya.A.,  
 Kurochkin.I.V., Lee.Y., Lenhard.B., Lyons.P.A., Maglott.D.R.,  
 Maltais.L., Marchionni.L., McKenzie.L., Miki.H., Nagashima.T.,  
 Numata.K., Okido.T., Pavan.W.J., Pertea.G., Pesole.G.,  
 Petrovsky.N., Pillai.R., Pontius.J.U., Qi.D., Ramachandran.S.,  
 Ravasi.T., Reed.J.C., Reed.D.J., Reid.J., Ring.B.Z., Ringwald.M.,  
 Sandelin.A., Schneider.C., Sempie.C.A., Setou.M., Shimada.K.,  
 Sultana.R., Takenaka.Y., Taylor.M.S., Teasdale.R.D., Tomita.M.,  
 Verardo.R., Wagner.L., Wahlestedt.C., Wang.Y., Watanabe.Y.,  
 Wells.C., Wilming.L.G., Wynshaw-Boris.A., Yanagisawa.M., Yang.I.,  
 Yang.L., Yuan.Z., Zavolan.M., Zhu.Y., Zimmer.A., Carninci.P.,  
 Hayatsu.N., Hirozane-Kishikawa.T., Konno.H., Nakamura.M.,  
 Sakazume.N., Sato.K., Shiraki.T., Waki.K., Kawai.J., Aizawa.K.,

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,  
 Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sakai,D., Shibata,K.,  
 Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
 Rogers,J., Birney,E. and Hayashizaki,Y.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)

12466851  
 22354683

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp/  
 URL:http://genome.gsc.riken.go.jp/

Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  
 Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,  
 Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,  
 Ohno,M., Sakai,K., Sakazume,N., Sakai,D., Sato,K., Shibata,K.,  
 Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and  
 Hayashizaki,Y. Direct Submission  
 Computational Analysis of Full-length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site (http://genome.gsc.riken.go.jp) for  
 further details.

## FEATURES

## source

1..333  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="DBA/2"  
 /db\_xref="taxon:10090"  
 /clone="1730080N15"  
 /cell\_line="CRL-1722 L5178Y-R"  
 /clone\_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 333;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY

211 CTGGTGGAGCGGCTGTCCGAG 231  
 |||||

## Db

102 CTGGTGGAGCGGCTGTCCGAG 122  
 |||||

## RESULT 11

## BY052633

## LOCUS

BY052633 335 bp mRNA linear EST 06-DEC-2002  
 BY052633 RIKEN full-length enriched, TIB-55 BB98 Mus musculus cDNA  
 clone 1730070L04 5', mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

BY052633 335 bp mRNA linear EST 06-DEC-2002  
 BY052633 RIKEN full-length enriched, TIB-55 BB98 Mus musculus cDNA  
 clone 1730070L04 5', mRNA sequence.  
 BY052633  
 VERSION  
 BY052633.1 GI:26158081  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 335)

**REFERENCE**  
**AUTHORS**  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, A., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierzki, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pallai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L. G., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

**COMPUTATIONAL ANALYSIS OF FULL-LENGTH MOUSE cDNAs COMPARED WITH HUMAN GENOME SEQUENCES** Mamm. Genome 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system—384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

**FEATURES**  
 source  
 Location/Qualifiers  
 1..335  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="BALB/c"  
 /db\_xref="taxon:10090"

**ORIGIN**

Query Match 4.2%; Score 21; DB 13; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 211 CTGGTGGAGCGGCTGTCCGAG 231  
 |||||  
 |||||

**Db** 111 CTGGTGGAGCGGCTGTCCGAG 131  
 |||||  
 |||||

**RESULT 12**  
**AA052685** 337 bp mRNA linear EST 13-SEP-1996  
**LOCUS** m070607.r1 Soares mouse p3NM19.5 Mus musculus cDNA clone  
**DEFINITION** IMAGS:404364 5', mRNA sequence.  
**ACCESSION** AA052685  
**VERSION** AA052685.1 GI:1543608  
**KEYWORDS** EST.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
**REFERENCE** 1 (bases 1 to 337)  
**AUTHORS** Marra, M., Haller, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

**TITLE** The WashU-HMI Mouse EST Project  
**JOURNAL** Unpublished (1996)  
**COMMENT** Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:248132  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 331.  
**FEATURES**  
 source  
 Location/Qualifiers  
 1..337  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:404364"  
 /dev\_stage="19.5 dpc total fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /clone\_lib="Soares mouse p3NM19.5"  
 /note="Vector: p77T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAATGGAGCGCGCGCTGTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p77T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

**ORIGIN**

Query Match 4.2%; Score 21; DB 9; Length 337;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 211 CTGGTGGAGCGGCTGTCCGAG 231

```

|||||
54 CTGGTGGAGCGGCTGTCGAG 74

Db
RESULT 13
BB869636 338 bp mRNA linear EST 27-NOV-2001
LOCUS BB869636 RIKEN full-length enriched, pooled tissues, intestinal
DEFINITION BB869636 Mus musculus cDNA clone G630014K09 5', mRNA sequence.
ACCSSION BB869636
VERSION BB869636.1 GI:17115846
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 338)
Akimura,T., Arawaka,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akashira,S.,
Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
prepare full-length cDNA libraries for cap-trapper-selected cDNAs to
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Iehikawa,T., Ozawa,K., Tanaka,T.,
Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
source
location/Qualifiers
1. 338
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="G630014K09"
/clone_lib="RIKEN full-length enriched, pooled tissues,
intestinal mucosa, etc."
/note="pooled tissues ; (tissue type=intestinal mucosa,
dev_stage=adult, sex=male), (tissue type=accessory
axillary lymph node, dev_stage=adult, sex=male),
(tissue type=jejunal and colic lymph node,
dev_stage=adult, sex=male), (tissue type=gall bladder,
dev_stage=adult, sex=male), (tissue_type=vesicular gland,
dev_stage=adult, sex=male)"

dev_stage=adult, sex=male), (tissue_type=spinal cord,
dev_stage=11 days embryo), (tissue_type=brain,
dev_stage=13 days embryo), (tissue_type=spinal cord,
dev_stage=13 days embryo), (tissue_type=lung, dev_stage=14
days embryo), (tissue_type=brain, dev_stage=15 days
embryo), (tissue_type=ovary and uterus, dev_stage=10 days
pregnant adult, sex=female), (tissue_type=cortex,
dev_stage=0 day neonate), (tissue_type=cerebellum,
dev_stage=1 month neonate), (tissue_type=diencephalon,
dev_stage=16 days neonate, sex=male), (tissue_type=medulla
oblongata, dev_stage=16 days neonate, sex=male),
(tissue_type=cerebellum, dev_stage=21 days neonate),
(tissue_type=testis, dev_stage=8 days neonate, sex=male)"

ORIGIN
Query Match 4.2%; Score 21; DB 10; Length 338;
Best Local Similarity 100.0%; Pred.No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 211 CTGGTGGAGCGGCTGTCGAG 231
|||||
Db 108 CTGGTGGAGCGGCTGTCGAG 128

RESULT 14
BB859603 345 bp mRNA linear EST 29-SEP-2000
LOCUS BB859603
DEFINITION BB859603 UI-M-AP0-abe-c-05-0-UI.r1 NIH_BMAP_MST Mus musculus cDNA clone
ACCSSION BB859603
VERSION BB859603.1 GI:10375692
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 345)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE 97044477
PUBMED 8889548
COMMENT
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse.
FEATURES
location/Qualifiers
1. 345
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AP0-abe-c-05-0-UI"
/dev_stage="27-32 days"
/lab_host="DHI0B (Life Technologies)"
/clone_lib="NIH BMAP MST"
/note="Vector: pMT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_MST library is a non-normalized library
constructed from mouse striatum. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
```

by Bonaldo, Lennon and Soares, Genome Research 6:  
791-806, 1996. Tissue provided by Ms. Annie Novakovich,  
Zivic-Willer Laboratories."

## ORIGIN

Query Match 4.2%; Score 21; DB 10; Length 345;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGTGGAGCGGCTGTCGAG 231  
|||||  
DB 89 CTGTGGAGCGGCTGTCGAG 109

RESULT 15  
BY168519  
LOCUS  
DEFINITION  
Mus musculus CDNA clone I830072B19 5', mRNA sequence.  
BY168519  
EST.  
Mus musculus (house mouse)

REFERENCE  
AUTHORS  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 366)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Otsu, N., Saito, R., Suzuki, H., Yamanaka, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,  
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,  
Ravasi, T., Reed, J.C., Reid, J., Ring, B.Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Iehi, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gs.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
Miyazaki, A., Murata, N., Nakamura, M., Nomura, K., Numazaki, K.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by David A. Hume ( Depts. of Biochemistry  
and Microbiology/Parasitology Institute for Molecular Bioscience  
University of Queensland Brisbane, Q 4072 Australia ) whose  
assistance we gratefully acknowledge.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details

## FEATURES

## source

Location/Qualifiers  
1..366  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="I830072B19"  
/tissue\_type="bone marrow"  
/cell\_type="macrophage"  
/clone\_lib="RIKEN full-length enriched, bone marrow  
macrophage"

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 366;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGTGGAGCGGCTGTCGAG 231  
|||||  
DB 54 CTGTGGAGCGGCTGTCGAG 74

RESULT 16  
BY078228  
LOCUS  
DEFINITION  
RIKEN full-length enriched, pooled tissues, adult spleen,  
etc. Mus musculus CDNA clone K630013G08 5', mRNA sequence.  
BY078228  
EST.  
Mus musculus (house mouse)

REFERENCE  
AUTHORS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 371)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Otsu, N., Saito, R., Suzuki, H., Yamanaka, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,  
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,

Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Akawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

CONTACT: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Aizawa, K., Akimura, T., Arakawa, T., Itoh, M., Kawai, J., Konno, H., Hirozane, T., Imotani, K., Ishii, Y., Nomura, K., Numasaki, R., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numasaki, R., Ono, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

Location/Qualifiers

1. 371  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone\_lib="RIKEN full-length enriched, pooled tissues, adult spleen, etc."  
/note="(dev\_stage=adult, tissue\_type=spleen, sex=male), (dev\_stage=adult, tissue\_type=kidney, sex=male), (dev\_stage=adult, tissue\_type=testis, sex=male), (dev\_stage=adult, tissue\_type=thymus, sex=male), (dev\_stage=adult, tissue\_type=heart, sex=male), (dev\_stage=adult, tissue\_type=colon, sex=male), (dev\_stage=adult, tissue\_type=stomach, sex=male), (dev\_stage=adult, tissue\_type=livver, sex=male), (dev\_stage=13 days embryo, tissue\_type=whole body, sex=mix), (dev\_stage=14 days embryo, tissue\_type=whole body, sex=mix), (dev\_stage=16 days embryo, tissue\_type=whole body, sex=mix), (dev\_stage=17 days embryo, tissue\_type=whole body, sex=mix), (dev\_stage=15 days pregnant, adult, tissue\_type=amion, sex=female), (dev\_stage=10 days neonate, tissue\_type=brain, sex=mix), (dev\_stage=10 days

neonate, tissue\_type=thymus, sex=mix), (dev\_stage=10 days neonate, tissue\_type=heart, sex=mix)"

ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 371;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 0;

QY 211 CTGGTGGAGCGGCTGTCGGAG 231  
|||||  
Db 114 CTGGTGGAGCGGCTGTCGGAG 134  
|||||

RESULT 17  
BB871048

LOCUS

DEFINITION

BB871048 RIKEN full-length enriched, 13 days embryo spinal cord Mus musculus cDNA clone G630026J17 5', mRNA sequence.

VERSION

BB871048.1 GI:17117258

KEYWORDS

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1 (Bases 1 to 372)  
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hirozane, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, K., Okido, T., Saito, R., Saito, K., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

TITLE

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.  
e mouse tissues.

Location/Qualifiers

1. 372  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"

FEATURES

source

```

/clone="G630026J17"
/tissue type="spinal cord"
/dev stage="13 days embryo"
/clone_lib="RIKEN full-length enriched, 13 days embryo
spinal cord"

ORIGIN
Query Match          4.2%; Score 21; DB 10; Length 372;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231
|||||
Db 74 CTGGTGGAGCGGCTGTCCGAG 94
|||||

RESULT 18
W48899
LOCUS          384 bp mRNA linear EST 28-MAY-1996
DEFINITION    mc88c10.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:355602 5', mRNA sequence.
ACCESSION     W48899
VERSION       W48899.1 GI:1337363
KEYWORDS      Mus musculus (house mouse)
SOURCE        Mus musculus
ORGANISM      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 384)
AUTHORS       Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE         The WashU-HMI Mouse EST Project
JOURNAL       Unpublished (1996)
COMMENT       Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:227402
Seq primer: 5'Primer
High quality sequence stop: 356.
Location/Qualifiers
1..384
/organism="Mus musculus"
/mol type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:355602"
/sex="unknown"
/tissue type="embryo"
/dev stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCATCTGAATGGAGCGGCGGGAATTTTGTGTGTGTGTGTGTGTGT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p773 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

ORIGIN

```

```

Query Match          4.2%; Score 21; DB 14; Length 384;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231
|||||
Db 104 CTGGTGGAGCGGCTGTCCGAG 124
|||||

RESULT 19
AA914086
LOCUS          396 bp mRNA linear EST 14-APR-1998
DEFINITION    vy92a03.r1 Soares mammary_gland NbMMG Mus musculus cDNA clone
IMAGE:1313644 5', mRNA sequence.
ACCESSION     AA914086
VERSION       AA914086.1 GI:3053478
KEYWORDS      Mus musculus (house mouse)
SOURCE        Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 396)
AUTHORS       Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE         The WashU-HMI Mouse EST Project
JOURNAL       Unpublished (1996)
COMMENT       Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:683940
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 384.
Location/Qualifiers
1..396
/organism="Mus musculus"
/mol type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1313644"
/sex="male"
/tissue type="mammary gland"
/dev stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares mammary_gland NbMMG"
/note="Organ: mammary gland; Vector: p773D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTTACCATCTGAATGGAGCGGCGGGAATTTTGTGTGTGTGTGTGTGTGT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

ORIGIN
Query Match          4.2%; Score 21; DB 9; Length 396;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231
|||||

```







Tissues were provided by Hiromitsu Nakauchi ( Dept. of Immunology  
Institute of Basic Medical Sciences University of Tsukuba 1-1-1  
Tennodai,Tsukuba,Ibaraki 305-8578 ) whose assistance we gratefully  
acknowledge. Please visit our web site  
(<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

1. 421  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="K1C0003B04"  
/tissue\_type="bone marrow"  
/cell\_type="mast cells"  
/clone\_lib="RIKEN full-length enriched, bone marrow mast  
cells"

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 421;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 0;

QY 211 CTGTGGAGCGGCTGTCCGAG 231  
|||||  
Db 63 CTGTGGAGCGGCTGTCCGAG 83  
|||||

## RESULT 22

AA020238

LOCUS

DEFINITION mh49g09.r1 Soares mouse placenta 4NBMP13.5 14.5 Mus musculus cDNA  
clone IMAGE:445888 5', mRNA sequence.

ACCESSION

AA020238

VERSION

AA020238.1

KEYWORDS

EST.

SOURCE

Mus musculus

ORGANISM

REFERENCE

AUTHORS

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE

JOURNAL

COMMENT

The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:271224  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 439.

## FEATURES

source

1. 448  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:445888"  
/sex="unknown"  
/tissue\_type="placenta"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse placenta 4NBMP13.5 14.5"  
/notes="Organ: placenta; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5',  
TGTTACCAATCTGAAGTCGGAGCGCGGAAATTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M.Fatima Bonaldo."

## ORIGIN

Query Match 4.2%; Score 21; DB 9; Length 448;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGTGGAGCGGCTGTCCGAG 231  
|||||  
Db 94 CTGTGGAGCGGCTGTCCGAG 114  
|||||

## RESULT 23

CA600107

LOCUS

DEFINITION wawic.pk006.k4 wawic Triticum aestivum cDNA clone wawic.pk006.k4 5',  
end, mRNA sequence.

ACCESSION

CA600107

VERSION

CA600107.1

KEYWORDS

EST.

SOURCE

Triticum aestivum (bread wheat)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2002)  
Contact: Scott V. Tingey  
Crop Genetics  
E. I. DuPont de Nemours and Company  
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
Tel: 302-631-2602  
Fax: 302-631-2607  
Email: Scott.V.Tingey@USA.dupont.com  
Seq primer: M13.

## FEATURES

source

1. 489  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="wawic.pk006.k4"  
/tissue\_type="anthers"  
/lab\_host="DH10B"  
/clone\_lib="wawic"  
/notes="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; To  
examine gene expression patterns in wheat anthers  
undergoing meiosis. Library was made at the Waite  
Institute in Australia."

## ORIGIN

Query Match 4.2%; Score 21; DB 14; Length 489;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 GAGGCGGAGCGGCTGTGGCC 456  
|||||  
Db 188 GAGGCGGAGCGGCTGTGGCC 208  
|||||

## RESULT 24

AA790556

LOCUS	AA790556	501 bp	mRNA	linear	EST 06-FEB-1998
DEFINITION	vml7h08.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:1244127 5', mRNA sequence.				
ACCESSION	AA790556				
VERSION	AA790556.1	GI:2850676			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 501)				
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.				
TITLE	The WashU-HHMI Mouse EST Project				
JOURNAL	Unpublished (1996)				
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LINML ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. MGI:657815 Seq primer: -28ml3 rev2 ET from Amersham High quality sequence stop: 500. Location/Qualifiers 1..501 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:1244127" /sex="male" /tissue_type="mammary gland" /dev_stage="4 weeks" /lab_host="DH10B" /clone_lib="Soares_mammary_gland_NbMMG" /note="Organ: mammary gland; Vector: pT7TD-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAATGGAGCGCGCAATGTTTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."				
FEATURES					
source					
ORIGIN					
Query Match	4.2%	Score 21;	DB 9;	Length 501;	
Best Local Similarity	100.0%;	Pred. No. 3.5e+02;			
Matches	21;	Conservative	0;	Mismatches	0;
				Indels	0;
Gaps					0;
Qy	211	CTGGTGAGCGGTGTCCGAG	231		
Dbb	44	CTGGTGAGCGGTGTCCGAG	64		
RESULT 25					
BF581564					
LOCUS	BF581564	544 bp	mRNA	linear	EST 12-DEC-2000
DEFINITION	602101087F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4224572 5', RNA sequence.				
ACCESSION	BF581564				
VERSION	BF581564.1	GI:11652576			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				

High quality sequence stop: 506.

## FEATURES

Location/Qualifiers

1..548

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clones="IMAGE:1195764"

/sex="male"

/tissue\_type="mammary gland"

/dev\_stages="4 weeks"

/lab\_host="DH10B"

/clone\_lib="Soares mammary gland NBMNG"

/note="Organ: mammary gland; Vector: pTV7T3D-Pac

(Pharmacia) with a modified polylinker; Site 1: Not I;

Site 2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer [5,

TGTTACCAATCTGAAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTT

T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pTV7T3 vector.

RNA provided by Dr. Minoru Ko, Wayne State Univ. Library

constructed and normalized by Bento Soares and M. Fatima

Bonaldo."

## ORIGIN

Query Match 4.2%; Score 21; DB 9; Length 548;

Best Local Similarity 100.0%; Pred. No. 3.5e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGTTGGAGCGGCTGTCGCG 231

|||||

Db 69 CTGTTGGAGCGGCTGTCGCG 89

## RESULT 27

BM070219

LOCUS

DEFINITION

musculus cDNA clone IMAGE:5669363 5', mRNA sequence.

ACCESSION

BM070219

VERSION

BM070219.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 565)

Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,

Lemishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,

Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,

Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,

Cardenas,M., Gibbons,M., McCann,R., Cole,R., Teagareishvili,R.,

Williams,T., Jackson,Y., and Bowers,Y.

Endocrine Pancreas Consortium

Unpublished (2000)

CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:

Washington University Genome Sequencing Center For information on

obtaining a clone please contact: Juliana Brown

(brownjfas.harvard.edu)

MGI:1955689 This sequence now available from the IMAGE consortium,

for clone orders contact: info@image.llnl.gov

High quality sequence stop: 435.

Location/Qualifiers

1..565

FEATURES

source

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="ICR"

/db\_xref="taxon:10090"

/clone="IMAGE:5669363"

/sex="Both for embryonic & newborn, male for adult and

adult islet"

/dev stage="Embryonic day 10.5, E12.5, E16.5, newborn,

adult, mixed"

/lab host="DH10B"

/clone lib="Melton Normalized Mixed Mouse Pancreas 1

NI-MMS1"

/note="Vector: pSPORT1; Site 1: Not I; Site 2: Sal I; Five

libraries representing E10.5/12.5 pancreatic bud, E16.5

pancreas, newborn pancreas, adult pancreas, and adult

islets of Langerhans were separately constructed using

SuperScript Plasmid Library kit (Life Technologies). cDNA

was made by oligo-dT priming and size-selected by column

fractionation. Libraries were amplified once on solid

support and plasmid DNA from each library was prepared

and mixed in equal amounts. The mixed library DNA was

normalized by method #4 from Bonaldo, Lennon, and Soares

1996 Genome Research 6:791-806; 0.5 microgram

single-stranded mixed library plasmid DNA was mixed with

5 micrograms PCR product representing mixed library

inserts and hybridized to an EcoT of 6. Single-stranded

(unhybridized) plasmids were isolated by hydroxyapatite

chromatography and used to make this library."

## ORIGIN

Query Match 4.2%; Score 21; DB 12; Length 565;

Best Local Similarity 100.0%; Pred. No. 3.6e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGTTGGAGCGGCTGTCGCG 231

|||||

Db 84 CTGTTGGAGCGGCTGTCGCG 104

## RESULT 28

BU582780

LOCUS

DEFINITION

spermatoocytes Mus musculus cDNA clone IMAGE:6366860 5', mRNA

sequence.

ACCESSION

BU582780

VERSION

BU582780.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 573)

McCarrey,J., Eddy,M., Marra,M., Hillier,L., Clifton,S., Pape,D.,

Martin,J., Wylie,T., Dante,M., Bowers,Y., Theising,B., Gibbons,M.,

Ritter,E., Teagareishvili,R., Ronko,I., Maguire,L., Kennedy,S.,

Bennett,J., Waterston,R. and Wilson,R.

NIHES Mouse

Unpublished (2002)

CONTACT: McCarrey/Eddy NIHES Mouse

NIHES Mouse

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed and donated by J. McCarrey, Ph.D. (Southwest

Foundation for Biomedical Research, Dept. of Genetics) - excision

done by E.M. Eddy, Ph.D. (National Institutes of Health, National

Institute of Environmental Health Sciences).

MGI:2044292

Seq primer: -40RP from Gibco

High quality sequence stop: 423.

FEATURES  
source

Location/Qualifiers  
1..573  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6366860"  
/sex="male"  
/note="Organ: testis; Vector: pBluescript SK+ (Stratagene); Site 1: EcoRI; Site 2: XhoI; cDNA oligo df-primed [5'-(GA)10-ACTAGTCGAGTCTTTT-3'] and directionally cloned using 5' linkers 5'-AATTGGCAGGAG-3' and 5'-CTCGTGGCG-3'. Size selection of >400bp material gives average insert size ranging from 1-2 kb. Library was mass excised (from lambda-UniZAP-XR) and resulting single-stranded phagemids were prepped and transformed into DH10B. Library constructed and donated by J. McCarrey, Ph.D. (Southwest Foundation for Biomedical Research, Dept. of Genetics); excision done by E.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences)."

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 573;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231

|||||  
Db 19 CTGGTGGAGCGGCTGTCCGAG 39

## RESULT 29

AA265232 591 bp mRNA linear EST 20-MAR-1997  
LOCUS m248f08.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA  
DEFINITION clone IMAGE:716679 5', mRNA sequence.  
ACCESSION AA265232  
VERSION AA265232.1 GI:1901502  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 591)  
REFERENCE Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
CONTACT: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:442175  
Seq primer: -28mJ3 rev2 ET from Amersham  
High quality sequence stop: 506.

FEATURES  
source

Location/Qualifiers  
1..591  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"

/clone="IMAGE:716679"  
/sex="mixed"  
/tissue\_type="pooled organs"  
/dev\_stage="7 day"  
/lab\_host="DH10B"  
/clone\_lib="Barstead mouse pooled organs MPLRB4"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTGCAATTCGGTACCl, digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed by Bob Barstead."

## ORIGIN

Query Match 4.2%; Score 21; DB 9; Length 591;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231

|||||  
Db 31 CTGGTGGAGCGGCTGTCCGAG 51

## RESULT 30

CE288123 627 bp DNA linear GSS 26-SEP-2003  
LOCUS tigr-gss-dog-17000333823656 Dog Library Canis familiaris genomic,  
DEFINITION genomic survey sequence.  
ACCESSION CE288123  
VERSION CE288123.1 GI:36060388  
KEYWORDS GSS.  
SOURCE Canis familiaris (dog)  
ORGANISM Canis familiaris

## REFERENCE

AUTHORS 1 (bases 1 to 627)  
Kirkness, E.F., Rafna, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and Venter, J.C.  
The dog genome: survey sequencing and comparative analysis  
Science 301 (5641), 1898-1903 (2003)  
22875432  
14512627  
Contact: Kirkness EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.

## FEATURES

## source

Location/Qualifiers  
1..627  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="Dog Library"  
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

## ORIGIN

Query Match 4.2%; Score 21; DB 29; Length 627;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 GCTGGCCCCCGACGCGCGCTG 152

|||||  
Db 318 GCTGGCCCCCGACGCGCGCTG 338



prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5']  
 GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTTNN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5']  
 GAGAGAGAGAGTCCAGGCTCAATTAATTAATPARACCCGCCCC 3'.  
 cDNA was cleaved with XhoI and SstI."

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 670;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 CTGCTGGAGCGGCTGTCGAG 231  
 |||||  
 Db 55 CTGCTGGAGCGGCTGTCGAG 75

## RESULT 33

BU757970/c  
 LOCUS 673 bp mRNA linear EST 10-OCT-2002  
 DEFINITION UI-1-CF0-apb-d-09-0-UI.s1 NCI CGAP\_P1Tr1 Mus musculus cDNA clone

## ACCESSION

VERSION BU757970.1 GI:23720040  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 673)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

## AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

## JOURNAL

Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Dr. Minoru Ko, Dr. Janet Rossent, Dr. Janet Rossent

## COMMENT

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
 Seg primer: M13 FORWARD

## FEATURES

Location/Qualifiers  
 1..673  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-1-CF0-apb-d-09-0-UI"  
 /tissue\_type="Trophoblast"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP\_P1Tr1"  
 /note="Organ: Placenta; Vector: pTTT3-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
 NCI CGAP P1Tr1 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTTT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The

sequence tags for this library are GTGTG, ATCAT, GGGTG, GGTGTG. For additional information, contact: Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
 TAG TISSUE=placenta mouse  
 TAG\_LIB=UI-1-CF0  
 TAG\_SEQ=GGTGTG"

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 673;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 CTGCTGGAGCGGCTGTCGAG 231  
 |||||  
 Db 650 CTGCTGGAGCGGCTGTCGAG 630

## RESULT 34

AK002881  
 LOCUS 686 bp mRNA linear HTC 20-SEP-2003  
 DEFINITION Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610041D15 product:hypothetical protein, full insert sequence.

## ACCESSION

VERSION AK002881.1 GI:12833190  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)

## JOURNAL

MDLINE 99279253  
 PUBMED 10349636  
 REFERENCE 2

## AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

## TITLE

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

## JOURNAL

MDLINE 20499374  
 PUBMED 11042159  
 REFERENCE 3

## AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

## TITLE

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

## JOURNAL

MDLINE 20530913  
 PUBMED 11076861  
 REFERENCE 4

## AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

## TITLE

Functional annotation of a full-length mouse cDNA collection

## JOURNAL

MDLINE Nature 409, 685-690 (2001)  
 REFERENCE 5

## AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

## TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

## JOURNAL

MDLINE Nature 420, 563-573 (2002)  
 REFERENCE 6

## AUTHORS

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirakawa, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,

Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toyota, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

**TITLE**  
Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT**  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

**FEATURES**  
source  
1..686  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:0610041D15"  
/db\_xref="MGI:1897448"  
/db\_xref="taxon:10090"  
/clone="0610041D15"  
/sex="male"  
/tissue\_type="kidney"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
c1..434  
/note="unnamed protein product; hypothetical protein (evidence: truncated-longest-ORF)  
putative"  
/codon\_start=3  
/protein\_id="BAB22428.1"  
/db\_xref="GI:12833191"  
/translation="ESHG1PAAAGGPAAGAAVRVAHPAGGAHSLRAATATA  
ALPRCSPPPARPRGSAACPAARREVQGHFRHTPTTPRPGPRAAGSGRQSVSTDM  
TIFGRGHVPVPSVIYTDLSHSSHPGRYAASNPNSDCPTDA"  
670..675  
/notes="putative"  
686  
/note="putative"

**polyA\_signal**  
211 CTGGTGGAGCGGCTGTCGAG 231  
|||||  
55 CTGGTGGAGCGGCTGTCGAG 75

**polyA\_site**  
687 bp mRNA linear EST 17-DEC-2002  
BY737644  
DEFINITION BY737644 RIKEN full-length enriched, TIB-55 BB88 Mus musculus cDNA

**Query Match**  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY**  
211 CTGGTGGAGCGGCTGTCGAG 231  
|||||  
55 CTGGTGGAGCGGCTGTCGAG 75

**Db**  
211 CTGGTGGAGCGGCTGTCGAG 231  
|||||  
55 CTGGTGGAGCGGCTGTCGAG 75

**RESULT 35**  
BY737644  
LOCUS  
DEFINITION BY737644 RIKEN full-length enriched, TIB-55 BB88 Mus musculus cDNA

# ACCESSION VERSION KEYWORDS SOURCE ORGANISM

## REFERENCE AUTHORS

clone I73008M03 5', mRNA sequence.

BY737644

GI:27150771

EST

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 687)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Nikaido, I., Oato, N., Saito, R., Suzuki, H., Yamanaka, I.,

Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Mogami, A.,

Schönbach, C., Gotohori, T., Baldarelli, R., Hill, D.P., Bult, C.,

Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,

Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,

Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,

Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,

Gariboldi, N., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,

Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,

Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,

Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,

Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,

Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,

Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,

Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,

Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,

Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,

Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,

Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,

Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, F.,

Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,

Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K.,

Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,

Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,

Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

1246851

Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp

URL: http://genome.gsc.riken.go.jp/

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,

Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,

Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,

Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,

Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,

Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,

Takeda, Y., Waki, K., Watabiki, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to



prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details

# FEATURES

source  
Location/Qualifiers  
1. .687  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="BALB/c"  
/db\_xref="taxon:10090"  
/clone="I73008M03"  
/cell\_line="TIB-55 BB88"  
/clone\_lib="RIKEN full-length enriched, TIB-55 BB88"

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 687;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 211 CTGGTGGAGCGGCTGCTCCGAG 231  
|||||  
Db 111 CTGGTGGAGCGGCTGCTCCGAG 131  
|||||

# RESULT 36

BE371234 601218984F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3587780 5', linear EST 21-JUL-2000  
LOCUS BE371234 698 bp mRNA  
DEFINITION BE371234.1 GI:9316597  
VERSION BE371234  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 698)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM8751 row: h column: 21  
High quality sequence stop: 626.  
Location/Qualifiers  
1. .698  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Czech II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3587780"  
/tissue\_type="spontaneous tumor, metastatic to mammary. Stem cell origin."  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Lu29"  
/note="Organ: lung; Vector: pCMV-SPT6T6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dn. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

# FEATURES

source  
Location/Qualifiers  
1. .698  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Czech II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3587780"  
/tissue\_type="spontaneous tumor, metastatic to mammary. Stem cell origin."  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Lu29"  
/note="Organ: lung; Vector: pCMV-SPT6T6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dn. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"  
Query Match 4.2%; Score 21; DB 10; Length 698;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 211 CTGGTGGAGCGGCTGCTCCGAG 231  
|||||

## Db

57 CTGGTGGAGCGGCTGCTCCGAG 77

RESULT 37  
BY709009  
LOCUS BY709009  
DEFINITION BY709009  
ACCESSION BY709009  
VERSION BY709009  
KEYWORDS BY709009  
SOURCE BY709009  
ORGANISM BY709009

REFERENCE  
AUTHORS

1 (bases 1 to 702)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyshew-Boris, A., Yanoisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: [genome-res@gs.riken.go.jp](mailto:genome-res@gs.riken.go.jp), [URL:http://genome.gsc.riken.go.jp/](mailto:URL:http://genome.gsc.riken.go.jp/)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

## TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: [genome-res@gs.riken.go.jp](mailto:genome-res@gs.riken.go.jp), [URL:http://genome.gsc.riken.go.jp/](mailto:URL:http://genome.gsc.riken.go.jp/)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers

1. .702

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="2310006P18"

/sex="male"

/tissue type="tongue"

/dev stage="adult"

/lab\_host="SOLR"

/clone\_lib="RIKEN full-length enriched, adult male tongue"

/note="Site 1: XhoI; Site 2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCGAGTCTCTTCGAGTCTCTTTTATTCCTCCCTCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCCGAGTCTCTTCGAGTCTCTTTTATTCCTCCCTCC 3']. cDNA was cleaved with XhoI and SstI."

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 702;

Best Local Similarity 100.0%; Pred. No. 3.6e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 211 CTGCTGGAGCGGCTGTCCGAG 231

|||||

DB 93 CTGCTGGAGCGGCTGTCCGAG 113

|||||

RESULT 38

BY710477

LOCUS

DEFINITION

musculus cDNA clone 2510049L02 5', mRNA sequence.

ACCESSION

BY710477

VERSION

KEYWORDS

EST.

BY710477.1 GI:27121700

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 714)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusica, V., Chatthia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godrik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G.,

Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Saito, R., Schneider, C., Sempie, C. A., Setou, M., Shimada, K., Sultana, R., Takanaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Varado, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, J., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp/

URL: <http://genome.gsc.riken.go.jp/>

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers

1. .714

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="2510049L02"

/tissue\_type="liver"

/dev\_stage="13 days embryo"

/lab\_host="SOLR"

/clone\_lib="RIKEN full-length enriched, 13 days embryo liver"

/note="Site 1: XhoI; Site 2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCGAGTCTCTTTTATTCCTCCCTCC 3']. cDNA was prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTAAATTAATCCCTCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI."

ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 714;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 211 CTGTGGAGCGGCTGTCGAG 231  
Db 84 CTGTGGAGCGGCTGTCGAG 104

RESULT 39

AK011122 715 bp mRNA linear HTC 20-SEP-2003  
LOCUS Mus musculus 13 days embryo liver cDNA, RIKEN full-length enriched  
DEFINITION library, clone:2510049L02 product:hypothetical protein, full insert  
sequence.  
AK011122.1 GI:12847042  
VERSION HTC; CAP trapper.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636

REFERENCE

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159

REFERENCE

3 Shibata, K., Itoh, M., Aizawa, K., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,  
Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861

REFERENCE

4 The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
PUBMED 11076861

REFERENCE

5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
PUBMED 12042159

REFERENCE

6 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,  
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,  
Furuno, M., Hanagaki, T., Harada, A., Hayatsu, N., Hiramoto, K.,  
Hirakawa, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,  
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,

TITLE  
JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216]  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.

COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTT 3'], cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. Second  
strand cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAGATTCGAGTAAATTAATCCCTCCCCCCCC 3']. cDNA was cleaved  
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.  
Host: SOLR.

FEATURES

Location/Qualifiers  
1..715  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM.DB:2510049L02"  
/db\_xref="MG1:1835853"  
/db\_xref="taxon:10090"  
/clone="2510049L02"  
/tissue\_type="liver"  
/clone\_libs="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="13 days embryo"  
c1..463  
/note="unnamed protein product; hypothetical protein  
(evidence: Truncated-Longest-ORF)  
putative"  
/codon\_start=2  
/protein\_id="BAB27415.1"  
/db\_xref="GI:12847043"  
/translation="GSERRVWRGSHGLPAAAGGPAAPAGGAARVAAHPAGGAA  
HSLRAATATAPRCSPPPAPRPSAARCPASPREVOGHERGTTPRPPGSPRAAEG  
SGRQSVTDWTLPGRHVFPSPVIYDLSHSSSPRYAASNFNSDCPTDA"  
699..704  
/note="putative"  
715  
/note="putative"

CDS

polyA\_signal  
polyA\_site  
715  
/note="putative"

ORIGIN

Query Match 4.2%; Score 21; DB 11; Length 715;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 211 CTGTGGAGCGGCTGTCGAG 231  
Db 84 CTGTGGAGCGGCTGTCGAG 104

RESULT 40

AK009191  
LOCUS Mus musculus adult male tongue cDNA, RIKEN full-length enriched  
DEFINITION library, clone:2310006P18 product:hypothetical protein, full insert  
sequence.  
ACCESSION AK009191



```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5352549"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

ORIGIN
Query Match          4.2%; Score 21; DB 12; Length 733;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 CTGGTGGAGCGGCTGTCCGAG 231
Db 78 CTGGTGGAGCGGCTGTCCGAG 98

BE282094          747 bp  mRNA  linear  EST 26-OCT-2000
601100912F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3493299 5',
mRNA sequence.
ACCESSION BE282094
VERSION BE282094.1 GI:9157425
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 747)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM8540 row: h column: 04
High quality sequence stop: 638.
Location/Qualifiers
1..747
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech II"
/db_xref="taxon:10090"
/clone="IMAGE:3493299"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Lu29"
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

ORIGIN
Query Match          4.2%; Score 21; DB 10; Length 747;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 211 CTGGTGGAGCGGCTGTCCGAG 231
Db 59 CTGGTGGAGCGGCTGTCCGAG 79

RESULT 43
BG976291
LOCUS
DEFINITION 602846558F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4976991 5',
mRNA sequence.
ACCESSION BG976291
VERSION BG976291.1 GI:14363928
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 761)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10971 row: d column: 16
High quality sequence stop: 681.
Location/Qualifiers
1..761
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4976991"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

ORIGIN
Query Match          4.2%; Score 21; DB 12; Length 761;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 CTGGTGGAGCGGCTGTCCGAG 231
Db 88 CTGGTGGAGCGGCTGTCCGAG 108

RESULT 44
BU962156
LOCUS
DEFINITION AGENCOURT_10614426 NIH_MGC_169 Mus musculus cDNA clone
IMAGE:6742852 5', mRNA sequence.
ACCESSION BU962156
VERSION BU962156.1 GI:24191740
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 826)
NIH-MGC http://mgc.nci.nih.gov/

```

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. Jonathan Kuo, NIMH  
CDNA Library Preparation: Michael Brownstein Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LNCM3081 row: j column: 03  
High quality sequence stop: 409.  
Location/Qualifiers  
1. .826  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6742852"  
/lab\_host="DH10B (T1-phage-resistant)"  
/clone\_lib="NIH\_MGC\_169"  
/note="Organ: Testis; Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggccgcctcggcc); CDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:  
5'-RAGCAGTGTATCAGCGAGAGTGGCATTACGCGGG-3' and  
5'-ATTCTAGCGCGAGCGGCCACATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH\_MGC Library."

FEATURES

source

ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 826;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231

Db 126 CTGGTGGAGCGGCTGTCCGAG 146

Search completed: June 16, 2004, 19:29:20  
Job time : 2108.33 secs

ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 826;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231

Db 49 CTGGTGGAGCGGCTGTCCGAG 69

RESULT 45

CB574912 854 bp mRNA linear EST 02-APR-2003  
LOCUS AGENCOURT\_12974065 NIH\_MGC\_136 Mus musculus CDNA clone  
DEFINITION IMAGE:30291619 5', mRNA sequence.

CB574912

CB574912.1 GI:29494442

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NIH-MGC <http://mgi.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. David Rowe

CDNA Library Preparation: Invitrogen Corp

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: NDAM334 row: c column: 20

High quality sequence start: 8

High quality sequence stop: 622.

Location/Qualifiers

1. .854

/organism="Mus musculus"

FEATURES

source

/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30291619"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_136"  
/note="Vector: pCMV-SF0R6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGCGGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 4.2%; Score 21; DB 14; Length 854;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231

Db 126 CTGGTGGAGCGGCTGTCCGAG 146

Search completed: June 16, 2004, 19:29:20  
Job time : 2108.33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:02:16 ; Search time 2074.33 Seconds  
(without alignments)  
10468.356 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543  
Perfect score: 501  
Sequence: 1 gcggtcccgagcgagca.....ccctcgggcagagagctga 501

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 15  
Total number of hits satisfying chosen parameters: 8604

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	501	100.0	53784	1	AMM223012	AJ223012 Amycolato
2	501	100.0	53789	6	A69720	A69720 Sequence 3
3	501	100.0	90445	1	AF040570	AF040570 Amycolato
C	31	6.2	82746	1	AF453501	AF453501 Actinosyn
	28	5.6	41097	1	AF016585	AF016585 Streptomy
	26	5.2	6567	1	AF521897	AF521897 Streptomy
	26	5.2	22131	1	AF506522	AF506522 Streptomy
7	26	5.2	132544	1	AF521085	AF521085 Streptomy
8	26	5.2	132544	1	AF521085	AF521085 Streptomy
9	24	4.8	78210	1	AB070949	AB070949 Streptomy
C	24	4.8	300425	1	AP005022	AP005022 Streptomy
	23	4.6	6378	6	AX751873	AX751873 Sequence
	23	4.6	18614	1	AE007033	AE007033 Mycobacte
	23	4.6	31422	6	E38021	E38021 Avermectin
12	23	4.6	31422	6	BD097650	BD097650 A method
13	23	4.6	31422	6	BD097650	BD097650 A method
14	23	4.6	31422	6	BD097650	BD097650 A method
15	23	4.6	64957	1	AB032367	AB032367 Streptomy
C	23	4.6	132544	1	AF521085	AF521085 Streptomy
	23	4.6	300050	1	EX248339	EX248339 Mycobacte
16	23	4.6	300050	1	EX248339	EX248339 Mycobacte
17	23	4.6	302675	1	AP005024	AP005024 Streptomy
C	23	4.6	302675	1	AP005024	AP005024 Streptomy
	23	4.6	347496	1	EX842577	EX842577 Mycobacte
19	23	4.6	347496	1	EX842577	EX842577 Mycobacte
20	22	4.4	723	8	AK063543	AK063543 Oryza sat
21	22	4.4	1019	8	AK107232	AK107232 Oryza sat
22	22	4.4	137312	2	AC135598	AC135598 Oryza sat
C	22	4.4	349737	1	EX572597	EX572597 Rhodopseu
	22	4.4	349737	1	EX572597	EX572597 Rhodopseu
24	21	4.2	727	1	AF144053	AF144053 Micromono
C	25	4.2	39314	1	SGR300302	SGR300302 Streptomy
	21	4.2	47981	6	AX112026	AX112026 Sequence
26	21	4.2	47981	6	AX112026	AX112026 Sequence
27	21	4.2	47981	6	AX112026	AX112026 Sequence
28	21	4.2	138203	1	AY310323	AY310323 Streptomy
C	29	20	3161	8	AK110400	AK110400 Oryza sat
	20	4.0	3835	8	AK112081	AK112081 Oryza sat
30	20	4.0	3835	8	AK112081	AK112081 Oryza sat
31	20	4.0	3900	1	AF262754	AF262754 Amycolato
32	20	4.0	4116	8	AK121779	AK121779 Oryza sat
33	20	4.0	7080	1	AF305325	AF305325 Comamonas
34	20	4.0	7853	1	MAU66560	MAU66560 Mycobacteri
35	20	4.0	12097	1	AE005116	AE005116 Halobacte
C	36	20	17589	1	AE004862	AE004862 Pseudomon
	37	20	20235	1	SERERYAB	SERERYAB
38	20	4.0	20235	6	AR049368	AR049368 Sequence
39	20	4.0	20235	6	AR095529	AR095529 Sequence
40	20	4.0	20444	1	SEERYABS	SEERYABS
C	41	20	31422	6	E38021	E38021 Avermectin
	20	4.0	31422	6	BD097650	BD097650 A method
42	20	4.0	31422	6	BD097650	BD097650 A method
C	43	20	41156	1	MSGY222	MSGY222
	20	4.0	43280	1	SFU78289	SFU78289 Streptomyce
44	20	4.0	43280	1	SFU78289	SFU78289 Streptomyce
45	20	4.0	48000	1	POL505006	AJ505006 Sorangium

ALIGNMENTS

RESULT 1  
AMM223012 53784 bp DNA linear BCT 09-FEB-1998  
LOCUS Amycolatopsis mediterranei genes encoding rifamycin polyketide  
DEFINITION syntheses, ORFs 1 to 5.  
ACCESSION AJ223012  
VERSION AJ223012.1 GI:2764760  
KEYWORDS ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.  
SOURCE Amycolatopsis mediterranei  
ORGANISM Amycolatopsis mediterranei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.  
REFERENCE 1  
AUTHORS Schupp,T., Toupet,C., Engel,N. and Goff,S.  
TITLE Cloning and sequence analysis of the putative rifamycin polyketide





[illegible]

QY	421	CGGGCCGCGACGAGCAGCAACTGATTCGACGCACCTGGAATCTCGGTCTCGTGCACGA	480
Db	15463	CGGGCCGCGACGAGCAGCAACTGATTCGACGCACCTGGAATCTCGGTCTCGTGCACGA	15522
QY	481	GCCTTCGGCGACGAGCTGA	501
Db	15523	GCCCTCGGCGACGAGCTGA	15543
 RESULT 2			
A69720			
LOCUS	A69720	Sequence 3 from Patent WO9807868.	
DEFINITION	A69720		
ACCESSION	A69720.1	GI:4774328	
VERSION			
KEYWORDS	unidentified		
SOURCE	unidentified		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 53789)		
AUTHORS	Schupp, T., Toupet, C. and Engel, N.		
TITLE	RIPAMICIN BIOSYNTHESIS GENE CLUSTER		
JOURNAL	PATENT: WO 9807868-A 3 26-FEB-1998;		
	CIBA GEIGY AG (CH)		
FEATURES	Location/Qualifiers		
source	1..53789		
	/organism="unidentified"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:32644"		
 ORIGIN			
Query Match	100.0%; Score 501; DB 6; Length 53789;		
Best Local Similarity	100.0%; Pred. No. 1.4e-220;		
Matches 501; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	GGGTTCCCAGCGCGAGCAGAACCGCATCTCTGCTGAAGCTGTCGCCGGCGCACGCGTTCACGAGGTC	60
Db	15043	GGGTTCCCAGCGCGAGCAGAACCGCATCTCTGCTGAAGCTGTCGCCGGCGCACGCGTTCACGAGGTC	15102
QY	61	ACGGTCTCGGCCACAGCGCGCGCGAAGGCATCGGCCCGCGCAGGCTTCACGAGGTC	120
Db	15103	ACGGTCTTCGGCCACAGCGCGCGCGAAGGCATCGGCCCGCGCAGGCTTCACGAGGTC	15162
QY	121	GGCTTCGACTCGTGCGCGCGGTCACCTCCGCAACGCTGCAACGCGGCCACC CGGCGCTG	180
Db	15163	GGCTTCGACTCGTGCGCGCGGTCACCTCCGCAACGCTGCAACGCGGCCACC CGGCGCTG	15222
QY	181	CGGTCGCCGCGAGCGCTGATCTTCGATACCCACCCCGGAGGCGCTGCTGGCTACCTG	240
Db	15223	CGGTCGCCGCGAGCGCTGATCTTCGATACCCACCCCGGAGGCGCTGCTGGCTACCTG	15282
QY	241	CGGTCGAACTCTTCGCGGAGGCGCGACGAGCGCTGCAACGCGGCCGGAAGACGACTCCGG	300
Db	15283	CGGTCGAACTCTTCGCGGAGGCGCGACGAGCGCTGCAACGCGGCCGGAAGACGACTCCGG	15342
QY	301	CGACTCTTCGCGCGCTGCGCTTCGCCCGGTTCAAGGAGGCGGCGCTCTCTGGACACGCTG	360
Db	15343	CGACTCTTCGCGCGCTGCGCTTCGCCCGGTTCAAGGAGGCGGCGCTCTCTGGACACGCTG	15402
QY	361	CTCGGCTTCGCGGACACCGGACCGGACCGGAGGCGGCGCGAGACCGAGCGGCGC	420
Db	15403	CTCGGCTTCGCGGACACCGGACCGGAGGCGGCGCGAGACCGAGCGGCGC	15462
QY	421	CCGGCGCGCGACGAGCAGAACTGATTCGACGCACTGGACATCTCCGCTCTCGTGCACGA	480
Db	15463	CCGGCGCGCGACGAGCAGAACTGATTCGACGCACTGGACATCTCCGCTCTCGTGCACGA	15522
QY	481	GCCCTCGGCGACGAGCTGA	501
Db	15523	GCCCTCGGCGACGAGCTGA	15543

RESULT 3  
AF040570 90445 bp DNA linear BCT 16-APR-2001  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295  
1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840  
1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934  
1935  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946  
1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989  
1990  
1991  
1992  
1993  
1994  
1995  
1996  
1997  
1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114  
2115  
2116  
2117  
2118  
2119  
2120  
2121  
2122  
2123  
2124  
2125  
2126  
2127  
2128  
2129  
2130  
2131  
2132  
2133  
2134  
2135  
2136  
2137  
2138  
2139  
2140  
2141  
2142  
2143  
2144  
2145  
2146  
2147  
2148  
2149  
2150  
2151  
2152  
2153  
2154  
2155  
2156  
2157  
2158  
2159  
2160  
2161  
2162  
2163  
2164  
2165  
2166  
2167  
2168  
2169  
2170  
2171  
2172  
2173  
2174  
2175  
2176  
2177  
2178  
2179  
2180  
2181  
2182  
2183  
2184  
2185  
2186  
2187  
2188  
2189  
2190  
2191  
2192  
2193  
2194  
2195  
2196  
2197  
2198  
2199  
2200  
2201  
2202  
2203  
2204  
2205  
2206  
2207  
2208  
2209  
2210

TATGLRLTDPAGQVPLTVESVRSPTFFVAEQGTTDALFRVDPWPIPLPTAETADEFLYE  
 ATSAEATLSALQAWLADFRVLTAVTGDCTGFECAAAIWGLVRSASQSEHRTVLADL  
 DDPAVLPAVVASQPOVNRNGVASVPLRTVTPROQADLPDEGTVLITGGTGTLCA  
 LTAHELTAVHGVRHLVAVSRGEAPELQBELTALGASVATAACVDARADLEAVLRAI  
 PAEPLPTAVITAGVITDGVVTPDRLATVRPKVDAAARLDELTREADLAAFLVL  
 SAAAGVLGNPCQAGTAANAELDALAQRSNLDLPVAVSIANGVWYATVSGMTEHLGDAD  
 LRLNORIGMSGLPATERGALLDAAITAGTTLAAKFDVAALRAKATAGAGGVFPLRLGL  
 ALPLPRAAKATASUTERGLAETEQAAALDVRRAAEVIGHSGASVHSRGTFRD  
 AGFOSLTAVEMRLAAATGLTSPAMIFDYKPKPALADHLRAKLGPSAANRAHEEDTH  
 AKADEPILVAMARCKPFGCVHSPLDMLRVADGADAVTEPADRGMDTLRLYEDPDH  
 EGTTVYFGAVFLDDAAGFDDAAFCGISNEPALMDPOORLLILETSMELFERAAIDPTTL  
 AQDILVPGVGNVDYDAGVAGSGFRLTGGSASVLSRVAVHYFGVSGPVTVDL  
 ACSSSLVALHVAQALQRGECSSMALAGVMWMTVETFEBSRQRCGLAPGRCKAPAD  
 GADTQSGEGVGLLLEKLSVRSQRHQHVLAVRGSASVNSQGSANGLUTAPNGSQSRV  
 IRKALAAAGLSTSDVDAVEATGCTTGLDPIEAEEALLATYQGNRETLFVSGVKSNGH  
 TQAAAGVAGVIGWVMARHGVLPRTLLEHVRPSSYVDVMSAGAVEILLTEARDVNSGHP  
 RRAGVSSFGITQTNHVLVEVAAPITTPQEPAEFLVPVLVSARTAAUGLRQAGRLA  
 AFLGDRTVRVPDAAYALATTAQLDHRVAVLASDRAQLCADLAAGFSGVVTTPVDG  
 KVLVLFTQCGQWAGMGRLEAETFPVDFPEAFEAACISHELAAAHVSGVLDLADAGELVA  
 DDTMTYTCQALFAVETALFVSGVRPCLLAGHSIGELAAAHVSGVLDLADAGELVA  
 ELAGRKQTRPALGAMVAQVATEDVAPILLDGTVCAVANGPDSVVLTSGLTSEAAVLVAD  
 BRAGORTRERLAVSHFSPILMEPLMDDFRAVARELITYRAGSLPVVTSGLTTEBLAAD  
 SPDYVGVGVRNAVFSDAATGALGASTFLELGGGALAAALGTGTEQFQSCVATLT  
 RKNGAEVLTALAEHLVRGVGDVMTLDEPATAVGTVLTPVAFQHQRFVMDVDTL  
 RAAVSTPPEAPIDVRPQVDVLVRESAAVLGHRDASFDLDRSPKDRHGFSLSAV  
 KMRRLRPTGVLEPSTLIPYPPAPVADHLRAELGERPAPVAPYTVRVSDEPTAI  
 VGNMRLTPGGASDPEELWLVAGRDPAVAGSFVDRGMDLDGLHPPDPAHAGTSYTRSG  
 GFLDHAQDFDGLFGISPREALMDPOORLLETLSWEALRAGVDPLUSARGSDVGVT  
 GITVHDYVTRLEVPDQGYTMTGTATSSVASGRVAVYFGPEPAVTVTATGSSSLVA  
 MHLAAQLRQECSSMALAGATWASPDALFEFSRQRLGADGRCKAASGADGTGWA  
 EGVGVVLERISVABERHVLAVLRGSANVDGASNGLTAPNGSVPOORVLRGALAGA  
 GLASDSDVVEGHGTGALDPIEVQALLATYGOERBOPLWLSKLSNGLHTQAAAGV  
 GAGTNNIMRWHRGNPATLHVDEETSVQVDSNGAEVLEITEAREWPTGRPRRAGVSSF  
 VAGVNTKHAHIIIEGPAAEASVEVSVAVPLVRSASGASLAGRAGRLAAVLENESLAG  
 VAGALVSGRATLNERAVYAGSREADQGLQALARGENAPGVTTGTACKPKVYVFPF  
 GOGSOWMGQRDLDDSPVFAARKECAAAEQMTDWSLDLRGDADLDRDLDVDPVQ  
 ASFAMGVLAASVTSPTDPAVLGSGIETAAACVSGASLDDAAKVAALRSQAIAG  
 ELAHRGMSVALSIDEDAVARLTPWANEVEAAVNSPSSVVIAGDAQALDEALAG  
 DLRVRRVASVDYASHTRIVBEALAEITLAKLADIDARVPAIPFYSTVLTGTMLEQAVDA  
 GYWNRLQVRFVSGVADLAGLGVTFVEIISAHPLVQVLPSEISDDAAVTVSGLRRDD  
 GGLRLLASAAELVYRVGVAATVPAAGWVDLPTAFADRRHFWLMEHAEVTAEEAGM  
 DGEVTLTAEQSDVLSALLELIVPQRGALSTVVPVLAQMDRRRSTAEKRLRYVT  
 WOPFLREAAQVPGGRWLAWPVAGTDLALKELTGQGLDILVRLEITEASRAQLBQLRN  
 VLAHDILTCVLSLALDGPADAAEITASTLALVQALGDTTTSAPLMLCTVINGAAEGL  
 QDAAVTAPACAAWMLGVAFAELDRWGLVDLPALDARTQAALIGVINGAAEGL  
 AVRRSGYRRRLVRAPVPEASATSRPWRGTVLTCGAEGLGRHSAWMLAQSASRLIV  
 TGTGDVELELTABELBFGTVPEFCDDTDRDADQLVADSEVTAVHAADIAQTSVVDT  
 GVADYEDVFAAKVQTAVMLDQITFEDTDLDAFVFSISIAGIWWGGGQGPAGAAVLLD  
 LIVEMRARGLKASTAWALDQIGIMDEAALAQRLRRGVTPMAPPLAVTAAVQVAG  
 NEKVAWMDLWAAITPAFTSVRPSLPADLPEAKAILRAAQDDEGDDGTASLADSL  
 RAPVDAQNRILILLKLVHRGASHVLYSGEAGIGPQAFQVBFQFDSLAANVLNRSHIAA  
 TGLRLPATLIDTPTPEALVGLYRLREADDGLDGRDDLRLVLAIVAPFRFEAG  
 VLDTLLGLADLTDETFGTDAETTEAAPAADDELIDALDISIGLVQRLGQTS"

misc\_feature  
 3883..5775  
 /gene="rifa"  
 /note="encodes polyketide synthase loading domain"

Query Match 100.0%; Score 501; DB 1; Length 90445;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-220;  
 Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGGTCCCGACGCCGACAGAACCCGATCCTGCTGAAGCTGGTCCGGGGCCACGCTTCG 60  
 Db 17590 GCGGTCCCGACGCCGACAGAACCGCATCCTGCTGAAGCTGGTCCGGGGCCACGCTTCG 17649

Qy 61 ACGGTGTCTGGCCACAGCGCGCCGAAGGCATCGGCCCGCCGACGCGTTCAGAGGTC 120  
 Db 17650 ACGGTGTCTGGCCACAGCGCGCCGAAGGCATCGGCCCGCCGACGCGTTCAGAGGTC 17709

Qy 121 GGCTTCGACTCGCTGGCGCGGCTCAACCTTCGCAACAGCCTGCACGCGCCACCGGGCTG 180  
 Db 17710 GGCTTCGACTCGCTGGCGCGGCTCAACCTTCGCAACAGCCTGCACGCGCCACCGGGCTG 17769

```

ARBWTVRADLGEQASPNDDVVFOVEAPVETSAKVTHVTYTAGAGGSAVQPSANVFN
PERGLYHTGDCDDKADLSADALRRYRTQGITLQMCVFLAERSPIAQTALDRQK
QDAAKVGKLVNGLVAYTASGSDAPKERVLAHLDQLAPLIRANADVTVVQAGFV
GAWGMYTKNGVAGVVSADQANKAVVDKLLTALPADRAVQLRTKFKKALGGG
STRI GHNDCLFSLADTDFQYDPAETPEYLAETRTLPMGGETCAVNPPTDCPTAT
AELSFFHWTYLDNQDVLRAWSAQGLAEVTAKLGYRFTLDRSDVFPTTTRGAALP
VRVNLNDGYAAVNORTVNLVRLNTTGTALTRPLTADPRTWPGARAIQAQAIQP
SLATQYALLLELADPALPGREYSIQTANTGLMDARTGLNDLKQTIIVS"
1825..2142
/ gene="asm36"
/ protein_id="AAM54114.1"
/ db_xref="GI:21449382"
/ translation="NVTASVLCARAGCGAPRAGATARWRCSTACKQAAVRLRRKA
GSASHEHEDEQLREALLAVARATLTTEATPPWSAQATALVGLARRVVELAAADAGV
SGG"
complement (2129..3307)
/ gene="asm35"
/ complement (2129..3307)
/ gene="asm35"
/ codon_start=1
/ transl_table=11
/ product="bicyclomycin resistant protein"
/ protein_id="AAM54113.1"
/ db_xref="GI:21449381"
/ translation="MAVGKADSAVGGWMLPALLIATVLTGTITIEDMSVPSPDIARELG
VGPAPQVLTNFTLGYCYLCHGSLRFGRRPVLGGSAVMLVGLGALCAPTIG
LLITAPVOGIGASAAVLAVFVIGDVRGPAALRMVGLVNAAMALMTCAPLGGAI
NRVAGRGNILAVCATALSLLVFLFLETRTRATATATASVLDYRLIGSRAYL
AASLTSLFLAFVLYFAASPELYTAAFGLVGTGALHLLVVASFAVSATGILP
LLGGPAGVVRGVLGAVVVGVLGGPVLGGVTFALFVCGFVAVCPVWFDMSV
FPELAGAASSATMGLRALVTLTLAGALSTGCPFPAPAAVWAGAAVAVLLSALSFR
"
complement (3450..4064)
/ gene="asm34"
/ complement (3450..4064)
/ gene="asm34"
/ codon_start=1
/ transl_table=11
/ product="transcriptional regulator"
/ protein_id="AAM54112.1"
/ db_xref="GI:21449380"
/ translation="MARNPERRAALLDAAVEVLAGEGAGLTFRAVDQAGVPAGTA
SNVFAGRAEITLIGASERVYQRLDQAPLGETPDGPRDRSRVTELMHALVDRVSAFTG
FLALLRLLEARNPELREVLITARVADLRFNAEHAASGLPGDGTVALLWALNWL
ILRLTLPLDITDQRRDLVATLVDRLLAQPEAGSPFDESCK"
4151..4723
/ gene="asm33"
/ 4151..4723
/ gene="asm33"
/ codon_start=1
/ transl_table=11
/ product="dihydrofolate reductase"
/ protein_id="AAM54111.1"
/ db_xref="GI:21449379"
/ translation="MRKLVYFIASIDGVIAGEGHDFTYTAEPWVHLRAEYPET
PTHVRPLVGMPLTPNRRYDVLVLMGRGSPALDAGVGSFYAHLRQVWVSRSLPADP
AVALVRADPVGLVRLKREDGLDWLCGADLAGQLLPEIDELVVKTYPPVAGGGTFL
FRAGAPARFPVDVAFDGGVVTYRPA"
complement (4854..5057)
/ gene="asm32"
/ complement (4854..5057)
/ gene="asm32"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAM54110.1"
/ db_xref="GI:21449378"
/ translation="MWITEQNARIPARRSSGRSNARRCAVWPRPQPSRATRAPGPRS

```

```

SANASRARSRRDRPTSLNRS"
5415..6461
/ gene="asm31"
5415..6461
/ gene="asm31"
/ codon_start=1
/ transl_table=11
/ product="RNA polymerase sigma factor"
/ protein_id="AAM54109.1"
/ db_xref="GI:21449377"
/ translation="WMIGYSADGRRRRPVRWGWVTEATWETDEDIKLAATLRDAEA
ANADSTRAYLKAIQRVRLLTADQVELARRVEAGLYAEHLDDLPPAPGSRPPVELR
RVYRDQKADKHLLEANLRLVVSIAKRYTGKMLLDVTQENGLLELRAVEKFPYAKG
FKFYATWIRQAITRALADQSRITRLPVHMEVVTKLARIERDLHRRGEATAEE
VAVELDVTPKEKVELRQHARTPVSIDQITGAEDTSGDLIADENAVAPLEAVSALL
SADLEAVLETLTPREAGVVRMYGFPDGVHLTLDDEVGRVYGVYTRERIRQIESRTWAKL
QHPAKSRGLGEGYLD"
complement (6501..9518)
/ gene="asm30"
complement (6501..9518)
/ gene="asm30"
/ codon_start=1
/ transl_table=11
/ product="cytochrome p450"
/ protein_id="AAM54108.1"
/ db_xref="GI:21449376"
/ translation="MVATGTRIPGPKPLPLVGNLLDVLTSDDTDVDFLDRCHREHG
IVALTIPAGORQVFASSHVELVAEMSCDSPSKAVHPALQVDRPAGDGLFTARGDENW
GKAHLDMFAPGTAMRDHFPANLDIAEQMLVWRFRPGDHRIDVDADMTLTLDTIA
LCAHARFNSFYDRADHPFDMVRLVSEAGERLERLPVQPLVGNORYRDDIATM
NRIDAVIYAARALPAGERPDDLRLMTCADPVTGERLSARNRYQLATFLIAGHET
TGLLSFAVRLHLAHPVLRKADAVDGLGDRVAFEDLARLDYLGQVLRFLRHP
TAPALAPDEPAELGGHAIGAGEPVLVMLPTLHRDPAVVRDPDVPDPERFAPERME
IPACWMPRGHARACIGRPEALQEAIVLVLORFDLADADPHRLTIKOTLTLPK
DILAVPRADPRADPGTATVTVVHQVPAITHRGTPLVHFGNSGSGEGELARTIAG
DGAAGWATSVAPLDDAVRALPASGPEVIVSSYNGAPPDPAHAFVRMLTODGPDLSG
VYLVLCGNLDSATYQRPVTLIDEAAMAAAGRLRERGATDARADFTDGMERWYEP
LWPLLSAECGVEGIGRFRVSDAADGIDLASAVVLENRELVRGPDAGSKRHL
LRPDGTSYRTGYSVLQNHDPDGVRAVARTVTRAVRVTVSSAPGTGLVPGVRA
RVDELITRCVDLSAPAGAVVARLAERCPCPERAELAAATGATLLELLELSPSCAVD
LALALPAPTRRLYSISSAAEQEAEVALTVSVTGVTSGYLSRVVRPGRVAVGIAS
PPSPFPADPTVPVVLIAAGTGIAFRGFLRARAALGEGPGEALLPFGCGPDLDD
YAEFAALGDMLEVDRAVSRHPDGEVVRHQHRLWQRDRVRVRELVDAGARYLGGDATR
VGPVAVELGRIGPGAGWLDLALRAGRYATDVF"
9664..10245
/ gene="asm29"
9664..10245
/ gene="asm29"
/ codon_start=1
/ transl_table=11
/ product="transcriptional regulator"
/ protein_id="AAM54107.1"
/ db_xref="GI:21449375"
/ translation="MSAAVRREQLVAALRVKMRDGLAAATTRAICAEADMPHGAHY
CFRSKQELYALTTLTGLNTSDAAMPATISDLPAVDNVRTMTAAFWTAFQADPEAEIV
LGEVLTAQREPELRELPKAWHQHREYVAKAVAHLDRLAEAEAGFDYSVDRLAELVSV
LDGVASSWFSHRDGLAKASLDQFAVMEALTR"
10393..10857
/ gene="asm28"
10393..10857
/ gene="asm28"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAM54106.1"
/ db_xref="GI:21449374"
/ translation="MTDTTTRHRTPVQNAALAVGAVFLLVGIAGFVPLGTLTYDGLF
AGHSDAALIGVFKVLAHLNHLVLAFCAGAGIAAGLAKAPRGARAFILGGGATYAVLW
YGVVVDHSTANFVPLNAADNWLHLGLAVGMIGLGAALGRNAHYRAARSDD"
complement (10912..11937)
/ gene="asm27"
complement (10912..11937)
/ gene="asm27"

```











EGPLGAMAWLWEQITGTMANTLSQADQARMSRGGLPIGVAEGMRLEFDLALRMPAL  
AVPIKLDLRADAAAGSVLALRLALVQTGRRQAQAGENGSLTARLASITPQBOE  
ELLDLVLTOAATVAGTYEISNVEAHAFNEAGFSLTSVELNRRLRATGLKLAATS  
VFDYTPPALARHLHGLKPAQSPGTGTHPLAELSKLEARLVETPVDDDSIRAQVITRL  
QVLLTTWSTANGTPTGDEALDFDAISDDDLFLDITTFSN"

CDS

9111..13136

/codon\_start=1  
/transl\_table=1  
/product="ShnB"  
/protein\_id="AAP45193.1"  
/db\_xref="GI:31211265"  
/translation="MLELBAEKSPVAIVGMACRFPDGVASPDQLWRLVSEGREGLSG  
FPEGKWDVEGFDADPXRKSYTDGRGFLHCAGLFDAGFGFISPREALMDPQOOL  
LLEVSMEALREIDPGLKSDGVVFSGVTTCGYGMRGETPGLGXLTGMXPXTFV  
SGRVSVFEGESVDITACSSIVAMHLAQALRGECSPMALAGGATGMANXTFV  
EFSROXXSXADQKXAYSADQGTWAGCAGVVVLELSEARNKHOLAVIRSSAVN  
QOAGNGLTAPNGPSQORVIRQALASAGLSAADVDVEHGGTGVLDPIEAQALLAT  
YGOEREPVLMGVSNGHGTQAAAGVVIKMWLAHGVNPAHLDPATLDPSSQVD  
WSAGAVELLTEARDPQTPRRAAVSSFSMGTNAHLIHSPEPQASVAPVSDV  
VPLVVSARSLIVGQARLASVGSASDVALPEVAAGLVSRRAVLSERAVVAGSRD  
QALSGALARGESHGVTGSLGAGVGRKVLVPPGQSPWGMGRLLCSAVFA  
ARVGCATLEWRWMSLVDLARGDCSVEFEREDVROPASFAVMVGLAAWESGVFA  
ADAVVHSGGEVAAACVSALSLIEDAVRVVAVRSQIISGVLGRGMAVGLSERAAV  
ARLQMDGRVEIGAVNPSVVISGTEALQGVLDALADGVSARMEVDFAHSRQV  
EAEELREAFADIRAQAPLIIFYSTVTGEWREAGVLDGGWYRNLRQVRFQFAIA  
DLIADGHTVFVSSSPALVWLINELADEANTVILVSGRDFGDLHRLFTSMADLF  
VGRIVDMGSLVLAAGADARRVDLPTVAFEHVWLPMTGSTTDAASLGSGADHPLFG  
LVIEPLVPAQXKRVQXAVGCPENGXRSVAVYSARDDBXGKTGXDXDQWTHAGTL  
TAAVPAQDFGTVPFPPGAERVSDFVFIEMAGRGYVYGFAGFLRGVWRREGEV  
FEVALPDPQHEGAEFGLHPALLDALQSLVLRPADGVDMRVPVFNGLRLHAGA  
SELVRTPSPGPDVSLQADGAGGVLVLSLIVARAVDVEQIDRMATDGRDALFEV  
DMSLPPAPASSVESIPPSALVASAEDVTDLADAAVPAVAVLEAVGCGDEHDLALTV  
RVLEVQWAFRAAGLAESRLVVVTRGAVFVGEGEVADPAGAAVGLVRAQAENPDR  
IVLLDLAADVDMGSLVPSXG"

CDS

13847..14533

/codon\_start=1  
/transl\_table=1  
/product="ShnB"  
/protein\_id="AAP45194.1"  
/db\_xref="GI:31211266"  
/translation="MPLPSPAKESKMTAEKLLSKLVTAELHDLRQQVARTDBPV  
AIVGMACRIPGGVGSQDRLWRLVSEGREGVSPPEGRGWDVEGLFDADPGRAGKSYTD  
RGFLHAGLFDAGFGFISPREALMDPQHWLRACGCTGPGCLPRWRCKSSWTG  
GCVLMORPAGVPRCCVLCIPVGGRRRPORTAVELPRGSPVSHRRSRKHCCSTWBGH  
RWSSSATSR"  
14963..15505  
/codon\_start=1  
/transl\_table=1  
/product="ShnB"  
/protein\_id="AAP45195.1"  
/db\_xref="GI:31211267"  
/translation="MPLPSPAKESKMTAEKLLSKLVTAELHDLRQQVARTDBPV  
AIVGMACRIPGGVGSQDRLWRLVSEGREGVSPPEGRGWDVEGLFDADPGRAGKSYTD  
RGFLHAGLFDAGFGFISPREALMDPQHWLRACGCTGPGCLPRWRCKSSWTG  
VSPITGLPPAVVSTPMAPR"  
15358..16812  
/note="polyketide synthase type I; PKS"  
/codon\_start=1  
/transl\_table=1  
/product="ShnA"  
/protein\_id="AAP63812.1"  
/db\_xref="GI:24637553"

CDS

14963..15505

/translation="MGVLADVHASGPOVALRSGAVLVRARVADTDRASTCRRLLDP  
DGTALITGTCALGALVARHLVVEHKIRSLVSRGPDAPGADLDDELTAALGARVR  
IVACIDRADRAGELIASVPRDAPLTAHVHTAGVGGVVTALPERLDVLRPRADA  
VAILDELTRHDLVAAFVLFSSAQTGNGPGGNLAASNAVLDALAVRRTAGIPATSV  
ANGVQDTGTSGLGVADQRRMARWGGLAASQGLELFDALAXDADVLAAXXNFA  
GLRAQASBPVHLLRGLVRAGRAAQASRREGLAGLQATPPVQREQLILDLVLR  
EVAALVGYSPKRVDPRAQDVGFTSVLAVELNRLAGLIGLRLTAFIDPTPRR  
MRRHLAELCEPGESEPADREIRALATTPLSRFRELGLMEQILHLVHVPGESAA  
APDTAEFKQDAGPLIAEMDIDNIVKRAKARKP"  
16813..22131

CDS

16813..22131

/codon\_start=1  
/transl\_table=1  
/product="ShnB"  
/protein\_id="AAP63813.1"  
/db\_xref="GI:24637553"  
/translation="MDARQNTETIAVASSKSVKEALRASLMENRLESEVQIRDSLT  
EPIAIVGMACRFPGGVSPPEELWELIADGRSAVEAFPTNRGWDLENLDPDRPGTT  
YVREGAFLHAGFEDAGFGFISQSETVMVMDPQOOLMLETSWAEFRAGIDPAAMRKN  
VGFAGMAAGQBYGTAFHSIPELEGYVMTGLASVLSGRVSYTFGFGPAPVITDTAC  
SSSLVALHMAQSLSESGESSLALVGTVMATPTAFVLTARAGLAKGRCKCAFAASA  
DGTNWAEGVULLERLSDAVNRGDEVILGVTRATVATVQDASNGLAAPNGPSQQRVIR  
QALAAAGLSPADVDIVEAHGTGTGADPIEAQALLTYGRNRPAGLPLWLGSKNLG  
HAGAAGVAGVITKMWAMRHGVLPTLHVDETPEDVMSAGAVELLTAHEWPEVGR  
RRAAGSGCAGCTNAHVLIEQATEPSTGNLPDEKARVLGDSVYVPLVVGARGKAGLAG  
AHLRGSFTORODTDLVIGOSLRSRGNLPDRAVLAADREALADLADVARAESAP  
GVVTFASRTGRTVLVFGQGTQWAGMALEASPVFAARMTSCAEVLDPITGWSL  
LDVVRGARSLEVDVLQFVSWALMYSLAALWEACGVVPAVVGHSIGEIAACYA  
GALSIDARLMVHRSLAEALVGRGNASLTADVKAVALIIEWPGFLEIAAVNGPA  
SVVVTGELPSEELLARCEADGIRARRINGAAHSSQIDVLHDSFVEALASVAGA  
SRVPLYSTVTRGLDHTTDFVHFRNRKQVQFPAIRSLVGDGHGVFIEVSAHPVL  
TSSVDVLADLEAGPAVVTGTRDDGGPRFLASLHLYTHGVKRVMEAVLGRGSEQ  
PVDLPTAFORORYLETAESRGDAPGLGVANHPILGAVTEIRGSDGVLTSLSL  
RTHPLADHAGAGVILLPGAAVELAVRAADEVGYLGVSELVIERPLVLPESGGVQR  
VMVGPDEGHRTVQVHRSREAGSGRWRHVSGLRVPEDQCAEFDLTQWPPPGATA  
VDPDAFAHDLAEAGYHYGPAFGQMRRAWTGVEVFAEVSLSPEAGADYGLHLPA  
LDDAAHTSLFRPDLSDSPKIALPFVWRDVLHADGASTLRVHLTLPADTIRHLA  
DTSQTPVASVDSLVLRPVVPELLRSGAAQDMFVAMEPLSVRSVDDDELKAVRTT  
AEDVRAAATAPRVLLDVGDRDTPDAARDLSGRVLEAVCAWLAEPAFQDTLLAL  
THISGAARVGDGPPVPLAVATAAGLLRAQSENVGRILLVDTGTTESARRLPDVLAAG

Query Match 5.2%; Score 26; DB 1; Length 22131;  
Best Local Similarity 100.0%; Pred. No. 0.25;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 TTCCAGGAGTCGGCTTCGACTCGCT 134

Db 2008 TTCCAGGAGTCGGCTTCGACTCGCT 2033

## RESULT 8

AF521085

LOCUS

DEFINITION

ACCESSTION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

gene

CDS

FEATURES

SOURCE

gene

CDS

FEATURES

SOURCE

gene

CDS

FEATURES

SOURCE

gene

CDS

FEATURES

SOURCE

gene

CDS

FEATURES

```
gene
CDS
/codon_start=1
/transl_table=11
/product="NanT1"
/protein_id="AAP42848.1"
/db_xref="GI:31044136"
/translation="MHDTTKDDPRASERKSGVRFAVSATLTGTALEWYDFAVYSAAAA
LVFGDLFFGSEDPDTGLTLLAFSYAVGVYVSRPLGGVFGRLGDIVIGRKVLIATLVLI
GAATFLIPLVTSTIGVAAPIALVLRPAQGVGVGEGWGAALLSEFGDSRRRGFY
ASNAQVGPAGNLANGVLAALGALLTEAQFEAWGRVAFLLSGALVAFGLMTAKLE
ETPVKAMEAQSRPAPREVTTPQRALLAILCRVGDVLIYAMFTVFVLIYARGE
LGMRSGLAALQVLQVLMPLAGAVSDRNRRVLYGCAAVAAAGVFFLFFPMIG
GGSWIFLAGVGVVHISLYGFAFIAEQSPRLRYTGSLSLAYTLAIGGAIAP
LFTFLISAYGSMPLALYIAVAANAAGVSLAGVLLGRDPEAAVDEDAQLATPAQAATA
ARTAV"
2064..2879
/gene="nanT2"
2064..2879
/gene="nanT2"
/notes="putative ABC transporter"
/codon_start=1
/transl_table=11
/product="NanT2"
/protein_id="AAP42849.1"
/db_xref="GI:31044137"
/translation="MLGHRPFGILLVYRSYAQALRKLTAEKSTTADRHNPNGSAPM
ARLTFELPDGSTRVDIVOLNAGYACRSQDDVAAHIAELAEIAGVTPSPVTPALYPA
PVLAAQIDRVAVOHRRTSGRAEWALVAGGELLTLTAACDHTDRDLGVHGVAKSNAG
PVLARARWLADVEPRDLTLRAWVTRDGTETEIOHGLAEILTPAYVVDVLRSD
ALPTGVTLISGTIPMTPGVDQFADTWKRVELGDPATGDTIRLAYDVHPMPPEPG"
3117..3818
/gene="nanT3"
3117..3818
/gene="nanT3"
/notes="putative two-component response regulator"
/codon_start=1
/transl_table=11
/product="NanT3"
/protein_id="AAP42850.1"
/db_xref="GI:31044138"
/translation="MSTRCTGSLTNGWVSVELGIPLSVVLVDHPVVRAGISAWCAA
DAPISVVAEGANYSHTGPGRGADVVVMDLLQNGRPAYDELQELVAQKRVVYTM
RDSQDAALTCDWLGSAITYTKAEQORHLVAIRAAEDIPYTPSLAGAFGSDTRQR
PVLSEVREVLVWFQSEKAVVAQSLGISERTVNTYLDVRVRIKYANAGRPAATKAKL
VARAVQGLIALDEL"
complement(3724..4923)
/gene="nanT4"
complement(3724..4923)
/gene="nanT4"
/notes="hypothetical chemoreceptor"
/codon_start=1
/transl_table=11
/product="NanT4"
/protein_id="AAP42851.1"
/db_xref="GI:31044139"
/translation="WGEDFROLITRHOIRAGGLVILVTLHIGSALPGLLGLIGEY
RHPVPLAAYGLLTILIGGSVVLGDRPWRGVPVPCACLTFFVASVATTSQVPADHY
FVSLHWSYSLACHWPTVLLGHRGPLVSGAELGHALVATALLTVDVPSRSGISMA
LSLSAGCFQNTIAAGKALLDSSAAIQOALRAOERVNRTIARVQIQADQRRYAEI
NATVPLLTGLASGLSDPEDDEVRHACALEARLURUPAESDCTFDPDLHEMRACIDV
AERNGTTVQVLRGDPLEPLPLRLALDPVIAAALAAQOTARTVTVVRGSDQVRGVV
VDALDRLPEPGKANGIRVTRVHAEDRLIVEACRIITARPPELSDRPEPWPPEWSWP
AGRSRT"
complement(4916..6106)
/gene="nanT5"
complement(4916..6106)
/gene="nanT5"
/notes="putative two-component sensor histidine kinase"
/codon_start=1
/transl_table=11
/product="NanT5"
/protein_id="AAP42852.1"
/db_xref="GI:31044140"
/translation="WSISLSFGIDPPASARLPGSAGPWAREYVAFNRRLALYGRAGV
```

```
MSCAVILVAAAMPGMGEMAPAAVVAALAWSWTHLRATTSAMRQRPPLALDITVMTA
LCLSORFTTPEAQTHAGTWLVSVFTAVAYQMOPPLTGILATLWLCQCAEAVGTAL
HPNLWSGVLRVIVVWTALARAVVRLVFSRRAADEAARARACQCEAAEARC
AAERHEAALHDFTACATLLIASAPWASRAETIRAOARDLLRLRAEQPVAVGVDLAA
ELLDHETAPHLRVINRFDVELGTTWEPRIAALRGAGEALRLVARVAGVDVATRAER
VDHIVVTIADSGVGFDPPEHPGHRIGIKRSIHGRMNWVGGRATVSRPGHGTIVRLE
WPRG"
complement(6768..7493)
/gene="nanR1"
complement(6768..7493)
/gene="nanR1"
/notes="putative regulatory protein"
/codon_start=1
/transl_table=11
/product="NanR1"
/protein_id="AAP42853.1"
/db_xref="GI:31044141"
/translation="MGLRVRVDEKGSPLRAQKLAIVALLIRSDOVVVAEQITREI
WGBLPPRRDTASIVICISLRKFLSRPGQNPITIRPGYMLRRASCECDFVQQLV
ADGRKCVRYGEPQAALCFEQALALWRGPAFNVQTGPIVAGFLKRLDEMRIECAEMS
AEAHLGRHRELVGRLSLVAEPFLRETFRQLMLALYRSQKSEALGIVESARRVL
DRELGLPGRKLRHLRHTVLAST"
complement(7573..8334)
/gene="nanR2"
complement(7573..8334)
/gene="nanR2"
/notes="putative regulatory protein"
/codon_start=1
/transl_table=11
/product="NanR2"
/protein_id="AAP42854.1"
/db_xref="GI:31044142"
/translation="MRVEILGRLVRVDRSSVTSISKIEVLLAALLVRADEVSSSED
LVKEIMGTAPROQVAALHYVYQSLRKLFRPNRPTESPIVTRBSGYVRLGDDDELDEH
VFQGLVROQRAAQAGRTHEACRAYEGLFWRGFPVLEDLDGAIINGFVFNWQAQARL
ECVSEVAGLEGRHREFVSYLYARIEELPNEAFRQLMLALYRSCRADALHYVQ
SARAVLSEELGEPSSAALKRLHHAIVLVADELEAA"
8919..17627
/gene="nanA1"
8919..17627
/gene="nanA1"
/notes="PKS"
/codon_start=1
/transl_table=11
/product="NanA1"
/protein_id="AAP42855.1"
/db_xref="GI:31044143"
/translation="MRFRYGVVMAGSSPSHAQQATSPVAIVGLACRLPGADPDPAFWR
LLRAGENAIVPVPDSRLTPFGSPPYFAGLLEHVDYTFDAGFFGISPREATMDPQORL
MLELAWEMEDAGLGKPKLAERTAVFTGAIWDDYATILHRRRPNDAITRHTWAGHL
RGLIANRVSHLLGLRGPSTLVDAAQSSGLVLAHLACSLRRGEADLALAGVNLILAE
ESMERAEAQFGLSPDGRCTEDARANGFVRGEGGMVLLKPLAAVADGDPVVCVIE
GSVNDGATDGLTRPSADAGTDVVRQARAAVSPAELOVELHGTGTGTPVGDRIEAA
ALGDALGDGDRCHRDADAPLRVSKTNVGHLEAAAGIALLLTALSIIHRRLLP
SLNFTNPPTGPIAELGLRVQTAFPWPDERRLTAGVSFGMGCTNCHVLAPEPPAP
AVLPDRTPAGDVSATAPVPMVWVWSAASPALTAQAAALYELRHAPGLHVDIIGH
ALATTRAFPHRAVVLGRDEDELVSRLDALASETQTCVIRGRAGGRVAFLEFSQGS
QRPQMGRLNAAVYFPADALREVCAHLDPMJTDITLDDLMFAEPDPEPIANETAY
TOPALFAIEVLRVTSWGVIPDLMHSHVGEITAAHVAGVLSIPDRLTVARGLR
MQSITAPGMAAWQATAEARQALEAGRGVLAANAPASVPSGDREAAVEATAAW
RARGKATVLSHAFHSPLDGLTLDLTVAGLTFAAPAIPIVSNITGGAAATQAL
RSPDYWADHARQAVRDAGVRLDAGVDVTFELGPDASTLCGMARESAANAGAPRP
VAVAVORRPEAQSFVSMAQAHVGVGVGDWMAAFAGHETRRAPLPTAFORDRHP
DGLDERGARPRSTSPVPSDREPVVVDASPADRAASFGELLALVTHAALVLGHNSP
DGIDPALTQKQGFDSLAATELSERLSAATDTELPALTIDPHFTPNVAAWMLRAHEG
QPTAAATAGTSMADPFAVAVVSCRYPGVSEGEALWRLVDSGVADGSEFGDGRGW
DLAELFRAPDGSGSATGCGFLYAGDFDAEFFGISPREALMDPQORILLISWE
LLERAGTPPASLAGSATVYGVATVDYVGRPLHEATDELGDHLLTGSTPSPVASGRVY
ALIGEGRALTVDTACSSSLVAMHLAAQALRQECDLALAGGVYMAVTPGMTFTSRQR
GLADPGRCKPFAAADGTGWSGEGAGLVLLEKLSDRNGHVQLAVIRGSVNOQGSN
GLSPNGFSQQRVIRQALNARLBFADVDVAEHGTGTTLGDPTEAQALLATYGGQT
DDRLWLGSIKSNIGHTQAAAAGVAGIKVNNALRHRLPALSHTDAPSPHIDWSQGT
RLLSPEVDWPGTDWPGSRPRAAVSFSFGISGTNAHLILEQAQDPHPEPTTSGVVP
```

WVLSARTADALRAQAAGLAETVAGAPRSPASPTSPASPADVGVMSLATTSTRADRRHRAV  
VSGTRDELLSGLFVADGSLAFAAASAGAAPGVVMPVPCQSGWGMVGLDSSPVF  
ARMAACEALLSGLFVADGSLAFAAASAGAAPGVVMPVPCQSGWGMVGLDSSPVF  
TTPAVVGHSGQGEIAAACVAGLGLSLRDCARVVALRSQALRALAGHGTMSIALSGAE  
RELADGAAARVAVVAVFNGPSTVVSGETDQVAAVVAACEAAGHARTIDVYASHG  
PQDLRLADRTDLDLSPGASDAVFSYAVTGARQTELDADYFNTLRQRFASA  
IDALLAAGVWFVTEVSGPHEVLPALRECFEEAEAAVPTLRLLQGGPDQVARLGD  
GFVAGLAVDWSWFVGDGEGHRRPTVELPTVFFORRYWLPADGREGREGRAG  
VCTPRAGHALLSASVELADGGLVLSGLPGLDGAAGVAGHVTAGVQVLPVGVADVWALLA  
ABAGASLEBELLRLAPLSGSLSEPSAGVLAQVAGADPESRRRLRISRRPAD  
AGAGEWTCHAVGSLAPGPPPAUTGTATVMPVPAAGALDPAGLVEAERRGYIG  
PALRGVVALWRGDADVADVALPEEAGCGGEGGADGDTAGFLHPVLVLDLALQPLRI  
ABPDGTGCGAGCEARLMLPFWMSGVRLWATGARAARVRLSLDGGGSDVADERLLI  
EVSPTGPAVSLVAVLPRVTVROVREASGAAGLFDLMTVPAQEPQSGAEDDAG  
CVAPLGAEPTEPCVDCRDTYDLPALLAALDAGAPLPSVMMWRPAAQPGCAEDAA  
LSAVRGVAALRAWAEPLRITVSLAVTVRGVAGAGSEGPVDLAAAANGCARGVQ  
AHPGDRIVLVDDVDMGADTIDIGAAAGLAAALGEPQVQLRGDTLLAPLARSA  
TEPGDVPFNGTVLTDGSGVLAGSVAEHLVRAEGVHLLVRFEGADGAYDTRDQ  
AQVMVTVDPRTDPAALERVVAQDPAPHTLVVHVAGLADTETSAGAGVAAVAGV  
RALHQATVAAALPSVRFTLSDAVWDGPAAPERAAGAFCAAVTDRRAGLGLDVA  
FGPWAADDGAGDSGRVTGLGADRGLALLRAACRADRPRLVADITRALTAHPA  
HELPAALRTLGASASAGRAPVRVRAAAPGRTTDMASRLVGLGPAERRRRLVLELV

Query Match 5.2%; Score 26; DB 1; Length 132544;  
Best Local Similarity 100.0%; Pred. No. 0.19;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GCCACCGGCTGCGGCTGCCCGGAC 194  
|||||  
Db 53661 GCCACCGGCTGCGGCTGCCCGGAC 53686  
|||||

RESULT 9  
AB070949  
LOCUS  
DEFINITION  
Streptomyces avermitilis polyene macroliide biosynthetic gene cluster.  
AB070949  
VERSION  
AB070949.1 GI:15824136  
KEYWORDS  
Streptomyces avermitilis  
Streptomyces avermitilis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE  
1  
Omura, S., Ikeda, H., Ichikawa, J., Hanamoto, A., Takahashi, C.,  
Shinoe, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,  
Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M.  
Genome sequence of an industrial microorganism Streptomyces  
avermitilis: deducing the ability of producing secondary  
metabolites  
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
21477403  
PUBMED  
11572948  
REFERENCE  
2 (bases 1 to 78210)  
Ikeda, H.  
Direct Submission  
Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of  
Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1  
Shirokane, Minato-ku, Tokyo 108-8641, Japan  
(E-mail: ikeda@c.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,  
Fax: +81-3-3444-6197)  
FEATURES  
Location/Qualifiers  
1..78210  
/organism="Streptomyces avermitilis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:33903"  
/note="Actinobacteridae"  
complement(1..1254)  
/codon\_start=1  
/transl\_table=1  
/product="putative lipoprotein"  
/protein\_id="BAB69301.1"

/db\_xref="GI:15824137"  
/translation="MGEALITLCITRSTRILATAGVLGGVLAIVTALGGTSNAATSGNGA  
GNPSTQARINQAGAEASDARIETTPQGTVSGTVNHPVSVTVTSGKLTVMTAVAV  
TGAELPGTSLDGTSTKNGRLERATKQIAAEAGDVKGRSTTGSATITTVGVANDFI  
GSLPESDSTGVAMPVTVNFKALSDKAQVSEIRVSSSSQORVUGVHNDHDFR  
PENYRPGSTVTVTLNRHGIQKTVFKRSQISTVDAKTKMTVRDCKTKITIPIS  
SGSPHPTNGQWVSEKFRQTHMGASVGLTEKNGKPSYDIKAVPHMRLTDSGTFI  
HGNVGTDSVFGCEANTSHCGVLDVRGDSHGQPAAPFFDHSIVGDVVIVKNSBEKT  
MLAPNGLSDNWMSWVAGSPAT"  
complement(1256..2290)  
/codon\_start=1  
/transl\_table=1  
/product="putative transposase"  
/protein\_id="BAB69302.1"  
/db\_xref="GI:15824138"  
/translation="MPVEFLSHDOVSRVGRFASAPTPGLEQFFRLDAKALRSASRKR  
APANRLGVAOVGCVRMGLGVFTEDLSVPEVVRFAAEQLGVDGEFAAGARQNR  
YEHAMEIRDAYIRRSAPAAQVRFLATRVWASWEGPRALDRAVGVLDNRVLDPG  
ITTLRLVAEVAQENALYRTLDEAVPEKQSMRDLKVPGRVRSLELRSPM  
GVSGADRLAFERRIMEAVMTAGPSTDIKELRVSVQSVQRVSLAGPRGK  
AVPQVSVLSLTPAALDEERAKGAPSWTITRYLHHPPIFQGVYAKWSQCLHSF  
GRSAQGYRDQ"  
2335..21081  
/gene="pceA1"  
2335..21081  
/gene="pceA1"  
/codon\_start=1  
/transl\_table=1  
/product="modular polyketide synthase"  
/protein\_id="BAB69303.1"  
/db\_xref="GI:15824139"  
/translation="MVSEELNATGRNDCSASADGIAVIGLACRLPGAADPAFAWR  
LI-SEGADAITVPPDRWDGAVADASAPRTGIRRGGLDSDIGDPAGFGLSPKEAA  
AMDQORLVELELAWEALDAHVAAATLRSTRTGTVFGVGMDDVATLHRSGLTASPH  
TVGLHRSIIANRVSYFLGNGPSLVDGQSSLSVHLACESLKSGESTIALAGV  
NUNVPSALGAAGLGLSPDRCFTFDARANGVVRGEGGIVVLKPLRAVADGTQVD  
YCVIRGSVNNDDGGDLTPQSGQEVRLAYERAGVNPADVPVLEHGTGKGV  
PIEAAALGAVLTGREPRLGVSAKTNVGLHGAAGITGLKAVLSLHRLPASL  
NFGTNPALDRLNLRVOTHESEMAEDGRPLLAGVSFGMGTTNCHIVLAGASL  
LVSHAPSPHMEPLAEFRNVAETTHAPRIPIVSNVTSELPSESPGTAAY  
ACEGATAGGGAAPTFWPLSAKTASGAQAQAAALLAHTAEPGLIALPDVGSLATGTA  
FEHRAVVGVRDDFLRELSTGIDDAITTRAGPRGELAFPLDSGQSQRAGWR  
ELASVPAFASALDEVCAHLDPLPRLEVPFAPVTGAEALDQTLTQTSLEAVE  
VALFRLESNDITPDVIMGHSIGELAAHVAGVLSLADACALVAARLMQALPAGGA  
MIAQAAEDEVRARIERTDRVSI AALNGPDSVVVSGDEDLATEIADAFAALGRKTSR  
LVSHAPSPHMEPLAEFRNVAETTHAPRIPIVSNVTSELPSESPGTAAY  
VREHRAVAFADQVETLELGPGLVSLASMAAGADESLFVALTRPE  
AQALLAAVSLHVGVEPDALFGGGRGARRKVALPTVAFERQWLDGDTSPVAT  
TGWAPVAPQDAIETDSSAPVGLRRLAGLAESEDEAVLDMVRAHIAAVLHGAER  
QVETWTFKDLGFLDSSLSLELNQSEATGLRPLSPGLLFDPHTPAALHARLAQSLGG  
GDAAAAPATVADPHEPIAIVAMSCRLPGVSSPEDLWRLLAGSGGDSIGFPTDRWD  
VETLYDPEPTGKSTRHGFLHEAAEFDAFGFISPREAAIDPQORLVLETAWA  
FERAGIDPATLRGRAGVFIGTAQDYGPRLHPADGVEGILLTGTTSVASGRVAYS  
FLEGPAVTVDTACSSVLVALLHVALQSLRSGECTMALAGGVTVMSFGMEVSRQSG  
LSADGRKAPAGADGTWAGVGLVLEKLSLDRNGHQVLAIVRGSVNVQDASNG  
DRTPNGSQORVVRQALAAAGAPADVAEHAHTGTGLDPIEAELIAAYGQGRDG  
LRLMGLSKNSNTGHTQAAAGVAGVIMVMNHGVLPTLTHVDEPTPHVDWSAGVR  
LTAVEWPARDLPRRAVSSFGISGTHNAHIVQOATSEEPARGTAVAGVPWVSA  
KSTVALREQAAGRLSLFLDAGVGVVRPEDVALSLATSRALERAAVVGGLDAELR  
QGLBALASGASGAAPYGVGVGFLPSQGSQRIGMGRRELYAAVPEAAVEVCA  
FLDAPVDVDEELNQTGTTPALPAVEVALFRLESNGRIPRTVAGHSVGEIAAAHV  
GVLUSLEDAKLVLSAQALMQALPDGAMVAQVATEDEVRHLTDGVGIAVAGNRSVV  
VSGAEDAVTAGVIFRAGRKTRSLKVSFAHPSFLMDPMLSEFAAVVGVTFEESHIP  
VSNLTKLAEVVTPEYVWVRVREAVFADGIELGLGVTFEVEVPGVGLAQAG  
CPDGDGTIVTVPLRADRPPEYALVTAIGRLHHTGISPQWQLFPGARRVLDLTPAFO  
RERYLDAAPRAATAGADAMDAEFWSVESEDRASIGALLGLEPTELDVVASLSAR  
KKRREQTIDGWRIRITWQPLDDMRAAPSGTWIYVYSEETAETAIAGLBELGVAL  
VPEAVGDETDRDALAFADAGQADGILFAAAQGEAGANLQRLVLLVQALDGAIG  
EAPLWCLTSGAVSTGTADLTDPVSAQLWGLGRIVALEQQRWGLVLPALDLSRAL  
ERLGLLAQADEQLAVRAGVYGRVLRAPLSAAPADAPWSPRGTVTGVTGALGG  
QVARMLAGAGAEHLVLTSRGPDAPGAELKLEELGAQVTVVACDAAORDALAEFL  
GRYIVNAVHTAGVLDGLVSLTPTERLNVHVRPKVDAALHLEHLETRDQDLAEVLF  
SSMTGVWGGGQAYGAANAFDLALAEHRRACQVPAALVAVGWSWGDGAAGDHL  
RRRGVIAIALPAISVLHGLHGTGETSMTVADVWDERFVPAFVGTFRPCPLLGVPEAR

EALEBARAAQSEVVPVVALVQRLGGLTGPQEQRIILLDVLVREQAATVLTGHTGKAFEA  
 DRTFRETFQSLTAVELHRLNTATGLRFTTLTDFDPTTALARHLGELLGRHGLE  
 TAAQAATAAAADPEALVAGMSRLPGVQSPDLWELSSGRDMSGFTTDRGWID  
 ALVDPDPDRPKYTSREGGFIQAGDRFDAGLFGISPREALAMPQORLLLETAWFAFE  
 RAGIALASVAPSTGVTGTINGODYANGLENAPETEGYALTCGAASVSGRLSYTFG  
 LEGPAAVTDVACSSSILVALHAAQAALSGECSMALVGGVTWTTTDPDLFVFSRQGLS  
 PDGRKAFAPAGTQWGBEGVLLVERLSDARRNGHOVLALVRGSAALNATQASNGIT  
 APNGQSQORVIRQALASAGITPADDAVEAHGTGRLGDIIEAQAALLATYQGHSDR  
 PVLWSKNSIHTQAAGVAGVTKVLMRHHGIPQTLHVEHGTTPAHVDSAGVRLIL  
 TEVPOVGTGPRRAVSSFGIGTNAHTIIEBAPAAAPAEAREHPPVPVMSAK  
 SEVALRAERLLAFATDDVSPVDAGFSSATTSALHRAAVIGSDPAELRAGLEALA  
 AGPAAVTVAGRAHSDAGLFGSGQSRGQREDIYAAPVPAFAIIDEVCAILLDAT  
 VVDVDELTGHTQSPALFAVFAVFLLESWGIRDYVAGHSVGEIAAARVAGVLSL  
 EDAKILVRAAALMKALPAGVMVAQAADEMLPHILDRAGAAINGPRSVVISGAE  
 DAVTAIEVPAQAQRTSLRKVSHAFSPILMDPMLERFAVVRGLTFSEWPIPVNSL  
 TGRLTSTYTPYVVRHREAVRADGVLTLHELGVTFVBEIGPGVLSALVQCGLDG  
 TDVTIIPALBAPPEQALTAQAQHVSGAEVDMHAFPPGARRVELPTVAQORRW  
 LDAPASAGDNRAVQGGAGHPLIGAANVPLADSVGHLLTGRLSKTHPWLVDHAGVVA  
 LPLGTAPELALITAAADAVGCDLLEELTLPVPPERGVALQVHVAGDGGSGRRPIT  
 VHSQIDGDYAHEDDSERAVRHSAGVLTATGPGVDSLTATPVPVGAEPVDDGPFY  
 DRLAGMALDYGPFQGLRAAWRGSDDFFAEVELPGAERTDAGAYGLHPALFDALHTV  
 WLGAVPEACTGNGLLPFAVSGVRLAAGASDLRVKVSRAAGTGTSTVSLVADGTGP  
 VASVDSLTLPVPADQLRAGSGSDDAVGLKMTVPVSLSPADCGWRLETVEDLAARAT  
 DDPADTEADAAAADAADVPCPTGTGADLATRVRETHAVLEVLQVWLAERFORLVL  
 VTRGDLAQSAVGLVRSAGTENPDRIVLVTSDDAHAGSGVDDADLARVLPGLIA  
 SGEPOAAVHADDEVPRVLRVAVPDAEPGLGSGTIVLLTGAAGSGGLGFARHLVA  
 EHGVRSLIVSRGGDAPGAELTADLTQAQADVTWAAACDVADRAAVRALLAGPGQSL  
 SATIHTAGVLDGGLISLTPERLDVFRPKVDAALNLHDIAGELSGDLSAFVLSVA  
 CILTGTCQQAANYAANTPLDALAHRRAQORATSLANGLWQDSSEASNTSGTDLDT  
 RIQRMGLAIPSSDGLRFLDALADRAVAPRLDGLGQGTVPVIRLGRVTRST  
 ARETAQTAQVLSQRLAGLAERAEQVLDLRTVEVAVLGHAGSGAGDADFSDI  
 GFDSLTAFLERNLRLAGNRLPATLIFDYNELTLAHFLTGEAAGSEAAATVPTTILS  
 TGTGTADEPTAIWMAKRPVGVQSPEDIMQVLFSGRDQVSGFPEGKRWDEOVLXDPN  
 PDQWGTSTYREGGLDADDAEFEGISPREALAMPDQORLLLETLSWEAFRAGIDP  
 ATVRSRTGVPAGVMWHDYGRVHTSPGLEGVLNCSAGSLASGSVSYTFEGEGRV  
 VTRGACSSULVALHAAQAALRSGECSMALVGGVTWMPGSPVFEFSRQGLSDGRCK  
 AFGADAGTQWAGVQGMMLVERLSDARRNGHGLVAVRGTAVNQDGSNGSLTAPNGS  
 LKSNIGHSQAAGGVGIVKMMAMQHGVLPTQLHVEPTPHVDMSAGAVOLLTEAVEM  
 PASDGPRAAVSFGVSGTNAHVIVEQAATSEVTAAGTAVGPVVPVLSAKSDAALRE  
 QAGRLISFLDAGVGGVGPEDVAVSLATSRVLECRANVVGEEIGEFGQGLEGVIS  
 GAAPVGVAGVRGFLPSGQSRIGMRILYAPVFAAAYDEVCAILLDVLVREQAATV  
 TLHQTGSTPALFAVAGVFLLESWGIRDYVAGHSVGEIAAARVAGVLSDEADLI  
 VASARALMQALPAGAMVAIQATEAEVLPHLTGVLGAAINGPRSVVLSGSDAVALI  
 AVEYPAQGRKTSRLKVSHPILMDPMLERFAVVRGLTFSEWPIPVNSLGRLEA  
 VYLPYVVRHREAVRADGVLTLHELGVTFVBEIGPGVLSALVQCGLDGHTHTVTV  
 PALRADRPEYALVTAVQGLTHGISEPQALPEGARRVLDPTYFAQORRWYLDAPQA  
 SGVRAUGAAGDHPLLGALATADSDGAVTLGSLGTHPMDHRAVNGILVLPGLIA  
 FVELAIRAGTDCGMITEDTLBELPLVPPERGVTIQAQVGPDESGRRQLSVHSRVD  
 DRSLRHATVLSAESAAPSAAKLGAWPAGAEAVDLTAFYEGMAAGLDTGPVFGQLR  
 SVWRGDVPAEVELPDSEAAEAFAHPALDHALAAGSLGLSDSPALPFAW  
 SGVVQOAGAAVVRKSLRTATGSLVTLAVDAGDGPVAVSESLSRAASAEOLPGAVG  
 GDQLRGLWTFRTLPSATDALMETVADFDALRESECTPDVVVPVVPVPCAGSGRETA  
 DRVHSTSEVALVQWMLAERAGRLAVTRPGDLAAHVAVGLVRSQSENPRIVLV  
 EASETEAVRLVLAAGSGPFAVRDGEVFPVRLVKATSEATGATGTPDFGVGVL  
 LTGASGALGVLAHLVAEHPGVSLLLSRRGAEPGAEVLEAEALWGAEVRWAACD  
 VADREAVSGLMFLVGEQSLVAGSVTAGVIGVLSITPERMEVREPKVDVNLHVE  
 TDRMDGLAAMPVSSVAGVAGQAYAAANGSFLDAPFSAHRRREGPLATSLAWGWE  
 QSGMVTGLVADRAERAGSVGLPLPEEGLRFLDALASDEAVLAPVRLDTGNARAG  
 EAPPVLRVPAARAAARTARAAPASASLAELEENPEAREKTVLIDVIRAVAGL  
 GHASDTRVREPAQDGLGFSLTAVELNRLNKATKATLPAVTFDHTPALLARHVF  
 MAAAGAEAPALVDSLADLRVQELVTLTAAEARDRIILSQGLLAIAGEFGKAPA  
 QEGFGTGGSDLESATDDEVFDLLKKEFGIS"

gene

CDS

/gene="pteA2"

/gene="pteA2"

/codon\_start=1

/transl\_table=11

/product="modular polyketide synthase"

/protein\_id="BAB69304.1"

/db\_xref="GI:15824140"

/translation="MSCRYFPGVDSPEDLMELVAGGRDAVSEFPADRPAGWLDLALYDPD

PERGTSYAREGAFILDLARFDAGLFGISPREALAMPQORLLLETLSWELEFAGIDA  
 ASLAGRTGVFAGVMHHEFLAALQNSQDLEGYLTGTGSGVAGRVAYTFLGEGPVA  
 TVDTACSSSILVALHAAQAALSGECSMALVGGVTWTTTDPDLFVFSRQGLS  
 PFAAGADGTGMBEGAGLLVERLSDARRNGHOVLALVRGSAALNATQASNGITAPNGS  
 QORVIRQALASAGITPADDAVEAHGTGRLGDIIEAQAALLATYQGHSDR  
 QORVIRQALASAGITPADDAVEAHGTGRLGDIIEAQAALLATYQGHSDR  
 IKSNIGHTQAAGVAGIKNVMMRHHGIPQTLHVEHGTTPAHVDSAGVRLIL  
 SASDQPRRAGISFGISGNTNAHTIIEBAPAAAPAEAREHPPVPVMSAK  
 APFAAGQPPVWISGTEEARAQRALLSRSTDAEPVLDVAFSLATRTTALBHRVL  
 QGRTVPAELAGLAVASGPARVARGLSTGGRVGLFSGQSGRGLMRELYASYP  
 AFAAAYDEVCAILLDATQADVAVDSELHRTGAQPALFAVEVALPRLLESWIRADYVAGH  
 SVGELAAHVAAGVLSLDDAKLYSARALMQALPAGAMVAIQATEAEVLPHLTGVL  
 IAINIGPRSVVVSAGDADTAIAEAFRQQRKSLKVSFAHSPILMDPILBEPVAV  
 RSUTHRPQIPVNSDNLRLAETPTQYVVRHREAVRADGVLTLHELGVTFVBEIG  
 PGVLSALVQCGLDGHTHTVTVPALRADRPEYALVTAVQGLTHGISEPQALPEGARRVLDPTYFAQORRWYLDAPQA  
 VDLPTAFQVRYWLDASPAAPGVRRAAGLSADHPLLGAAVSLAGGDERLLTGRLSL

Query Match 4.88; Score 24; DB 1; Length 78210;

Best Local Similarity 100.0%; Pred. No. 1.7; Mismatches 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 CTGCCCCGACGCTGATCTTCGAC 207

Db 60409 CTGCCCCGACGCTGATCTTCGAC 60432

RESULT 10

AP005022/c

LOCUS

Streptomyces avermitilis genomic DNA, complete genome, section

AP005022

AP005022 BA000030

VERSION

AP005022.1 GI:29603894

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS



and J. Ishikawa.

- \*1 Kitasato Institute for Life Sciences, Kitasato University
- \*2 National Institute of Infectious Diseases
- \*3 The Kitasato Institute
- \*4 National Institute of Technology and Evaluation
- \*5 School of Science, Kitasato University
- \*6 Institute of Medical Science, University of Tokyo
- \*7 RIKEN, Genomic Sciences Center

Following url is also available.

<http://avermitilis.ls.kitasato-u.ac.jp>.

#### FEATURES

##### source

1. 300425

/organism="Streptomyces avermitilis MA-4680"

/mol\_type="genomic DNA"

/strain="MA-4680"

/db\_xref="taxon:227882"

/notes="This strain is also named as strain: ATCC 31267,

NCIMB 12804 or NRRL 8165.-synonym: Streptomycetes

avermiticinus"

complement (382. .645)

/notes="SAV258"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAC67967.1"

/db\_xref="GI:29603895"

/translations="MNFALAPEVSVTDEHGTVLDDERTGRYLLNGTGALRALRL

DGTSPOGVASLQCRFPDAGTAQVEQVDNLLHALRTAKVAAS"

complement (642. .2552)

/notes="SAV259"

/codon\_start=1

/transl\_table=11

/product="putative membrane-associated Zn-dependent

protease"

/protein\_id="BAC67968.1"

/db\_xref="GI:29603896"

/translations="MFAVLPHLLILIVESTGMFELVLPDHPAGAAALAMPATAGPPP

TVIEHSGSPFIVGWNEEDVMVAVGHRVALLGRTTADADVLTRLRDVSVRDL

DGLAAAGSFHIVASNGEVRVQGLTSAACQIYYATGVTLAADRPQALADRIAD

IDEDQVALLSPFPPGVSRLWRCVSPPTGCLREIGRDTGRTROWMTAPDP

RLPGDMITQALAAVATCTAGQVVVSADLSGLDSTSCFLADADRDHDLVTVHY

QALGRSDRYWARQAEALPGARHLLVGGQAPDWPYAPLAHAVHREGLTVRSRA

TEHLAAAGSRRLQIGGDELFPSPMALHLVRHRLPPLAALPHLRALSKMRNR

PLTTARTLASPSYQWLAAACSNFPADRDWAGADWEPPLKLPMPATPHAAATVR

QLLAALATQPLATFPVQHEMLRLTRINGNATRNRSIHGATGVTFPHAPFTDRVLE

AILAVLEDRVALGQFKPLGTALRSIVPAGLLSRITKGDASAEIYAGLRRRGELLK

LMDDSHLVRMGFVDVLRVLRGLHADTRPLMDLSTLACELWLRLTSLPADTARATA

VPVPSPTVKEI"

3174. .3572

/notes="SAV260"

/codon\_start=1

/transl\_table=11

/product="putative transposase"

/protein\_id="BAC67969.1"

/db\_xref="GI:29603897"

/translations="MGRAGVAFITDEQRNNGYTFNQVDDGQIAGVFLDRDARRA

MTCGRARSQIAGVQLGTGVRFLGFLDNPDPAEYVAYVAEQIHPGSLVAGYGAER

TRWDHQIAIKDAYGYRDLKGDANKLRDGD"

3604. .3834

/notes="SAV261"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAC67970.1"

/db\_xref="GI:29603898"

/translations="MLGQGYVPLDLMIEGESQRMSERDALGLRALMWVIAIEVPV

YLVLVSVILTGQQLPRGTREEDGKPRATD"

4027. .4500

/notes="SAV262"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAC67971.1"

#### CDS

```
/db_xref="GI:29603899"
/translation="MQADAMAVAMELOQAAANRLIWEGPARGRTFLITLAFAGG
ARARITAGTQVQSLVGFQOAEELSRERVASKQTVAAVREPLAQAAATAAALMRCG
APAVVAATQLLNALADVENTARLEALEAFGRVAVNATPPRLSWARRRAVSSG"
complement (4811. .6445)
/notes="IS6 family; completely same as SAV/545 and SAV/5480
SAV263"
/codon_start=1
/transl_table=11
/product="putative IS6 family transposase"
/protein_id="BAC67972.1"
/db_xref="GI:29603900"
/translation="MILAECGELLMTMFPHLEGLVVEVSPGEGVLHIVARTVESVPV
PCDCATPSPRRHSQYRIADGAVGRQVSIETVTRLFCDQCARVTFAEQVDGL
TVYGRTRTQLRCLLSAIALAGAGERLAARLPVPSRTTILLALMALPDYATFP
RVLGVDEFATRKGHYGTVLVDCETHAPIDLLPDRESATFAAMLTHDPGVEIICRDG
GAFADGARTCAPDAQVADLWHLWHNLATVKSLSKSHSCLEPALSPEVILHP
PLSHAGRLAERARRHAAVHDLQGLTIRAVARLELSRNTVRRYARAVTWELATG
RWNLPSITLDPYKSYLHQHWEHTSGAKHLAELRERGTGSYSVVRDYLRRFRRTSG
DRPPARPFGRKVTGMITNPDNMNDQKQKAILARCPLEAATGTVRSFAAMWA
IRSASRLPEWIATARADEHGLRFPADGLIADLDADVLGLSTEWSSGCGVEGRVTDIKL
LKRQAGRAGLPLLRKRVLLVAADRRQRHVNTQTAH"
complement (6592. .8802)
/notes="SAV264"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC67973.1"
/db_xref="GI:29603901"
/translation="MTPATFPSPSTPPSPWQCGHGTTPEDPACGGRIVSGHTACLA
HUSDADRDAYLAGLSPGNVDHRTGTSFQQLDALLDALKRDPDTPGRUGVAFRSAT
FQDVGFRSATFOKNADFGSATFQGNAGFRLATFEGDAGFGSARFEGDAGFGSARFEG
DASFVSATFEGDASFVSATFEGDASFVSATFQSAALVGMWAAFGQDFKFSATFQGDVA
FGDAGFGDARFGLAARFARNARFGLARFGLASFMLVTGHGDAWFSATIPRGDASFVS
ATEQSDTDLEAATFSDHTRFDRATFQGDARFVEATFPGDTSFVSATLOSNANPLGVTF
ERDAFESSIFQGDATFVSATFECDAFCFQGVGFERSASLGPLVCAGRVKLSGAVFGCP
VTLSFATRCLECRTRWSSTAEURLRYATVDFAHVFEYFPLTIAASADPFVLEHGRPV
AEQSLDWRPIETVRIASLRGVAHLLSDVLSGCLFTGTVDLQDRIEGTCSFASV
PRETHRRWRPIETFORRALAERGFWRASOPTFVPGWNAVAGAHAGAPLAQALPVYRA
LRKAFDSNPEGAADFYEGEMEMRHDRTGSTRTERGLLHGYLLSGYGLASRALG
WLAAMLVITLLMGFLPKDSKPOVATCTVPGDGKVTLEIDKADPQNTGDRFTSE
REFKALTGLINSVVFSSSGQDITAGTYIEMASRVTEPVLGLLGLAVLARVRKX"
```

#### CDS

```
9294. .9707
/notes="SAV265"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC67974.1"
/db_xref="GI:29603902"
/translation="MGTRLRLRRADFSDCERKLTQKVFPVSRPGGVQGLITNTYLSP
TEYDILLASIPAAVLSKTRFSVPLGVDFDGLQGLVLAEEFTTDETSQTFPPSEC
VAEVTDDARFTGGRLVETSRQELLGLWLAEBYGLRPQ"
complement (9555. .10614)
/notes="SAV266"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC67975.1"
/db_xref="GI:29603903"
/translation="MHHVLRRESGCSVLNPLADANRLPVQFRPILLGNHRRRD
DNRFGWTPETLCRLSALTAAELSHAI PALAVIGEPARNLQSLSGQAVEVLRFP
ACRCRRMARGFHLGVRTTPPEAVCFQRHQLLGEQVRLHALPELQQANRRHCRILV
RCTRISALAREYAVARDCLRDWFQDGKFSALRQWNRKIDQLGEDPYGDETPVAPSP
G"
```

#### CDS

```
10830. .11180
/notes="SAV267"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC67976.1"
/db_xref="GI:29603904"
/translation="MLEDPAAHGVLDLCTWVHLHETGDEWPATRAHAEEFVLPHLRH
RVLVQVARASRLSLEITVDDSRQQRIVERGFWALWDEYESGGTVPQOQGIKCSLH
```

#### CDS

```
complement (9555. .10614)
/notes="SAV266"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC67974.1"
/db_xref="GI:29603902"
/translation="MGTRLRLRRADFSDCERKLTQKVFPVSRPGGVQGLITNTYLSP
TEYDILLASIPAAVLSKTRFSVPLGVDFDGLQGLVLAEEFTTDETSQTFPPSEC
VAEVTDDARFTGGRLVETSRQELLGLWLAEBYGLRPQ"
complement (9555. .10614)
/notes="SAV266"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC67975.1"
/db_xref="GI:29603903"
/translation="MHHVLRRESGCSVLNPLADANRLPVQFRPILLGNHRRRD
DNRFGWTPETLCRLSALTAAELSHAI PALAVIGEPARNLQSLSGQAVEVLRFP
ACRCRRMARGFHLGVRTTPPEAVCFQRHQLLGEQVRLHALPELQQANRRHCRILV
RCTRISALAREYAVARDCLRDWFQDGKFSALRQWNRKIDQLGEDPYGDETPVAPSP
G"
```

#### CDS

```
10830. .11180
/notes="SAV267"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC67976.1"
/db_xref="GI:29603904"
/translation="MLEDPAAHGVLDLCTWVHLHETGDEWPATRAHAEEFVLPHLRH
RVLVQVARASRLSLEITVDDSRQQRIVERGFWALWDEYESGGTVPQOQGIKCSLH
```







JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Chem. Biol. 10, 431-441 (2003)  
2 (bases 1 to 132544)  
Sun, Y., Zhou, X., Dong, H., Tu, G., Wang, M., Wang, B. and Deng, Z.  
Direct Submission  
Submitted (13-JUN-2002) Bio-X Life Science Research Center,  
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,  
China

## FEATURES

## Location/Qualifiers

1. .132544  
/organism="Streptomyces nanchangensis"  
/mol\_type="genomic DNA"  
/strain="NS3226"  
/db\_xref="taxon:204925"  
complement (682. .2052)  
/gene="nanT1"  
complement (682. .2052)  
/gene="nanT1"  
/note="putative integral membrane transport protein"  
/codon\_start=1  
/transl\_table=11  
/product="NanT1"  
/protein\_id="AAP42848.1"  
/db\_xref="GI:31044136"

## gene

## CDS

/translation="MDDTTKDDPRASRRSGVRRAFVASLTGTALDWDYFANYSAAAA  
LVFGDLFFPSEDPLTGTLLAFSTYAGVYSGRPLGGIVFGRIGDVI GRKKVLIATLVLI  
GAATFLIGLVYTYSTIGVAAPIALVLRFAQGVGGEWGGVALLSEFDSRRRGFY  
ASAAQVPPAGNLLANGALGALLTEAQQFEAWGWRVAFLLSGALVAFGLIRAKLE  
ETVFKMAEAQRPRAPIREVFTTOPRALAALICRVGPDVLYAMFTVFLVATGE  
LGMRSGLAANVLGSSLOVFLMPLAGVSDRINRVLYGCAVAGVWPEFLPEPMIG  
GGSWIPALGVVLGVFLHSLTIGPQAFIAEQFSPLRYTGGSLATYLAGIIGATAP  
LPTTLLSAYGSWLPIALYIAVAASVLAGVLIGRDEAAVSDAQATPAQAQTADA  
ARPTAV"  
/gene=".2879  
/gene="nanT2"  
2064. .2879  
/gene="nanT2"  
/note="putative ABC transporter"  
/codon\_start=1  
/transl\_table=11  
/product="NanT2"  
/protein\_id="AAP42849.1"  
/db\_xref="GI:31044137"

## gene

## CDS

/translation="MLGGRPSGILLVYRSYAQAKRKLTAEKSTTADRHNPNGSAPM  
ARLTFELPGDSTVMDIVQINAGYAGRSQDDVAAHIAELGALGVPTTPALYPVA  
PYLAQOQIDRVAVOHRRTSGAEAWALVAGDGLLITACDHTDRDLGVHGVAKNAG  
PDVLRARWLADVEPLDILTFAWVTRDGTETEOHGTIAELLTAYVWVLRSD  
ALPFGVLIISGTPMTFGVQFADTWVVELGDPATGTTIRLAYDVHWPPEIG"  
3117. .3818  
/gene="nanT3"  
3117. .3818  
/gene="nanT3"  
/note="putative two-component response regulator"  
/codon\_start=1  
/transl\_table=11  
/product="NanT3"  
/protein\_id="AAP42850.1"  
/db\_xref="GI:31044138"

## gene

## CDS

/translation="MSTRCTGSLNGWVSVELGPLSVLWDDHPVVRAGISAWCAAA  
DAPISVVAEGANVSVALHGPGRGADVVMLLLQNGRPAYDELQELVAQERKVVYTM  
RDSQDAALTCDMGSAITITKAEGRHLVKAIRAAABDIPTTPSLAGAFGSDTRQSR  
PVLVSREVLVEWFOSESKAVAQSLGISERTVNTYLDVRVRIKYNAGRPAATTKAKL  
VARAVQDLIALDEL"  
/complement (3724. .4923)  
/gene="nanT4"  
complement (3724. .4923)  
/gene="nanT4"  
/note="hypothetical chemoreceptor"  
/codon\_start=1  
/transl\_table=11  
/product="NanT4"  
/protein\_id="AAP42851.1"  
/db\_xref="GI:31044139"

## gene

## CDS

/translation="MGEDFRQLITRHQIRAGGLVILVTHIGSALPGLLGLGEY  
SLNPFATNPGLPLAELGLRVTAFGPPMDERRRLTAGVSGFMGGTGNCHVVVIAEPFAP  
RHPWIPLAYGLTLILGSGSVVLGDRPWPGRWPCALAGTFFVASVATTSQVPADHY  
FYSLWSYSLAGMTVLLGHRGPLVSGAFGLAHVATTAALLTVDVPSRSTGASVA  
LALSAGCFMTIAGAKILLDSAAIQALRAQERVTRIVARQIQADQRRRYAEL  
NATVPLLTCLASGLDPEDEVRHACAEARLRLEFABSDCTFPLVHEMAGVGVV  
AENGTTVOLAVRGDPLEPLPLRRALLDPIVIAALAAAGQATRVTVVRGSDQVRGVV  
VDALRRLPEPKANGIRVTVHAEDRLVVEACRITARPPELSDRPPPPWPPFWSNP  
AGRRSRT"  
complement (4916. .6106)  
/gene="nanT5"  
complement (4916. .6106)  
/gene="nanT5"  
/note="putative two-component sensor histidine kinase"  
/codon\_start=1  
/transl\_table=11  
/product="NanT5"  
/protein\_id="AAP42852.1"  
/db\_xref="GI:31044140"

## gene

## CDS

/translation="MSISISLPGCIDPPASARLPQSGAPWARYVAENRRLALYGRGV  
MSACVLGVAMPGEHAPAAVAAVLAWSWHLRLATTSAMRQPLLDLDTVMTA  
LCLSQRFTTPEAQTHAGTWILVSVSTAVAYQIMQPLTGTILATMLCCQADVGTAL  
HPNLEWSGVLSRVVWNTALARAIVRVLFESRAADEAARAAARQGEAAEARC  
AAREHLAALHDTACATLLIASAPWASMRATLRAQAADLLRLRAEQPVAGVYDLAA  
ELHDEIAAHLPLRVINRFDVELGTTWRPLAALRGGLGEALRNVARVAGVAVRTAER  
VDHVIWVTADSGVGFDPPEHPGHRIGIKRSIHGRMNWVGGRATVESRPGHGTIVRLE  
WPRG"  
complement (6768. .7493)  
/gene="nanR1"  
complement (6768. .7493)  
/gene="nanR1"  
/note="putative regulatory protein"  
/codon\_start=1  
/transl\_table=11  
/product="NanR1"  
/protein\_id="AAP42853.1"  
/db\_xref="GI:31044141"

## gene

## CDS

/translation="MSPLRVVDEKGGSPQAQKAAVLAVLLRSDOVVAEQLTREI  
WGSLPPRDTSIYVCISLRKFLSRPGQONPILITREGYMLRRRSECEDFVQQLV  
ADGRKCVYGEPAQALCFEQALAMRGPAGNVQTGPVAGFLKRLDEMRIBCAEWS  
AEAHLGRHRELVRSLRVAFVFFLETFTYQLMALYRSQKSEALGIYESARRVL  
DRELGLPGKRLRELHRTVLAST"  
complement (7573. .8334)  
/gene="nanR2"  
complement (7573. .8334)  
/gene="nanR2"  
/note="putative regulatory protein"  
/codon\_start=1  
/transl\_table=11  
/product="NanR2"  
/protein\_id="AAP42854.1"  
/db\_xref="GI:31044142"

## gene

## CDS

/translation="MRYEILGPLRVDRSSVTSISKRIEVLALALVRADEVVSSD  
LVKIMGTAPROAVAALHYVVSQIRKFLRRPNRTESPIVTARSGVYLRGLDDELDFH  
VFQGLVROGRAAORAGRTHEACAYEBGLRFPWGLVLELDGAI VNGFVCWAQOARL  
ECVSEFVAGLEGRHREFVSYLYARLEELPLNEAFYQLMALYRSGCRADALHVQ  
SARVLSEELGLEPFAALKRHHAVLVADELEAA"  
8919. .17627  
/gene="nanA1"  
8919. .17627  
/gene="nanA1"  
/note="PKS"  
/codon\_start=1  
/transl\_table=11  
/product="NanA1"  
/protein\_id="AAP42855.1"  
/db\_xref="GI:31044143"

## gene

## CDS

/translation="MRFYGVWAGSSPSHAQQTSPVAIVGLACRLPGAPDPFAFWR  
ILLRAGENAVVPDPSRLTFEFGSPPIFAGLEHVDITFDAGFFGLSPREATMPCQRL  
MLELAWEMEDAGLPKTLAERTAVTGAIWDDYATLLHRRRENDITATHTWAGLH  
RGLIANRVSHLLGLRPSLTVDAAQSSGLVAHLACELRRGEADALAGGVNII LAE  
ESMRARAQFQGLSPDCRCVTFDARANGFVRGEGGMVLLKPLAAAADGDPVVCVTE  
GSVANNAGATDGLTRPSADAQTDVVRQAQRAVSPAELVOYVELHGTGTGVPDPIEA  
ALGALGDADPGCHRDRAVDAPLRVGSVKINVGHELEAAAGIAGLKTALSIIHRRLLPP  
SLNPFATNPGLPLAELGLRVTAFGPPMDERRRLTAGVSGFMGGTGNCHVVVIAEPFAP

AVLPDRPTAGDVSARTAPVMPVWSAASPKALTAQAAALYEHLRAHPLGLHPVDLIG	
ALATTATPHRAAVRDELVSRLDADLASETQTSVIRGRAGGRVAPFLSGOGS	
ORPMGRELPAAYVEFADALREVCALHDDMLDDTDLDMFAEAPDGPGLNRTAY	
TOPALEVALVELYLTSMGVTDLHMGHSVGBLTAHVAVLSLDPDCTLVARGEL	
MQSITAPGAWAQATBAEQALEAYGGRVGLAAVNPASVVISGDRNAVAAATRAW	
RARGKATVLKSHAFHSHDGLDGLTRVAAGLTAPPAIPVNSNLTGAATBAOL	
RSPDYWDHARQAVREDAGVRLHCDAGVDTFLBGPDSLTTGMARSAAAWGADPRP	
VAVAVORGRPEAQSFVVSMAQARHGVGDVMAAFAGHETTRAPLTPTAFQDRHWP	
DGLDERCARPSPVPSVTRBPVVADPADRAASPGELLALVRTHAALVLGHNSP	
DGDIPALTFLQSLDAAETLSERLSAATTELPATLTEDHPTNPAVAWLAAHEG	
OPTAAPTATQSMADPDVAVAVSCVPGVSGESGALRLVDEGDVAVCEFPQDRGW	
DLAEFORAPGSGSATRGGLYAGDFDEAFEGISPREALAMDQQRLLLELSEW	
LLERAGTPASLAGSATGVVGTAVDYGPRLEHTEALDCHLLTGSTPVASGRVAY	
ALGLEGPALVTDTACSSSLVAMHQAQLRQECDCALAGGVTTMATPGTFEFSROR	
GLAPNGSPSORVROALANARLEPADVAVEAGHTCTTLGDPTEAQUALATYGGOT	
DLRPLWLSGSIKSNIGHTQAAGVAVLKNMALBHGRLPASLHIDAPSHLDSGTV	
RLIUSEVDPGTWPGSDRRRAAVSFGISGNTNAHLILEQAPDHPPEPTTSGGVVP	
WLSARTADALRAQAGLAEWVTAGAPRSPASPSPADVQWSLATTTSADRRHRAV	
VSGTDRDELLSGLRAVDAGLAPAAVSAGAAPGVMPVPGQSGWRGMVELLDSVPV	
AARMAACEAALGFDFVSLTVLRGAPGAPGPPSRVDVLQCLWAVMVSLAAVWESYGV	
TPTAVVCHSOGELAAACVAGLSLRDCGARVVALRSOALRALAGHTWASIALSGABAE	
REFLADIGAAARTVAVFNGPSTVVSPTDQVAAVVAACEAGHRARTIDVYASHG	
PQVRLADTIRTDLADUSPGASDAVYSVAVTGARQPTTELDADYFWTLNRPVRFASA	
IDALLAAGYRVFIEVSPHVLIPALRECFEEAEVAAATVTLRRDQGGPDQVAFALGD	
GVFAGLAVDSRWFVGDREAGDEHREFTVELTYPFQRRYWLAPDHGREGRTAG	
VGTFRPAGHLLSVAVELADGLVLSGRLPGDAWVGHTVAGVOLVPGAVLWDALLIA	
ADPAGASLEBELLRLAPLESGLSPGLSEPSAGVLAOVAVGAPSGERELRISRPAD	
AGAGEWCHAVGSLAGPGPAPADTGTATVFPVPPAGAEALDPAGLYERARRGYGV	
PALRGVVALVDGADILVADVALPEEAGGEGEGDGTAGTGLHPHLVLDLAAQOPALL	
AEPDGTGEGAGPEARLWLPFANSVGLRMATGAAPARVRLSPLDGGGDVADERELRI	
EVSDPTCAPLVSVALVRPRTVQVREASGAAGGLFALDWTVPAPQEPSGABDDAG	
CVALGEPATPEVPGDCRDTVDLPAALLADACAPLPSVVMRPPAADPGAPEDVQ	
LSAVRGVAALRAWVAEPRTVSLAVTVTRGAVAGGEGEPVDLAAAGACARGVQ	
AEHPDRVLVDVDDVMDADTDDICAAAGLAAALGEPQVALRGDILLAPLARSAA	
TPGCVGVPNGTVLVDVSGVLAGSVAEHLVRAGVRLHLLVREFGADGAYDTYDRQD	
AQHMVTVPRDTHAALERVVAQVDPHPLTVGVHVAGLSADIETSGAARWAVAGV	
RALHQAATAALPSVRSITDAATAGDGAAPERAAGACFAAATVDVRRRAGLGLDVA	
FGWAAADDDGGADGGRWTVGLCADRGALLRAACRADRPRLVAADIRTRALTAHPA	
HELPAALRTIGASASASAGGAPVRRVAAAPGRTTDSWLSVLGLGPAERRRVELV	

Query Match	4.6%;	Score 23;	DB 1;	Length 132544;
Best Local Similarity	100.0%;	Pred. No. 4.5;		
Matches	23;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

Qy	193	ACGCTGATCTTCGACTACCCAC	215
Db	95473	ACGCTGATCTTCGACTACCCAC	95451

AUTHORS  
TITLE  
JOURNAL

Garnier, T.  
Direct Submission  
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire  
Bacterienne Institut Pasteur 28.rue du Dr Roux 75724 PARIS cedex  
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the  
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary  
Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,  
Surrey KT15 3NP, UK. Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut  
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.  
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28  
rue du Docteur Roux, 75724 Paris Cedex 15, France

FEATURES  
source

1. 300050  
/organism="Mycobacterium bovis subsp. bovis AF2122/97"  
/mol\_type="genomic DNA"  
/strain="AF2122/97"  
/db\_xref="taxon:233413"  
165. 398  
/gene="secG"  
/locus\_tag="Mb1475"  
165. 398  
/gene="secG"  
/locus\_tag="Mb1475"  
165. 398  
/gene="secG"  
/locus\_tag="Mb1475"  
len: 77 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100% identity in 77 aa overlap). Probable protein-export  
membrane protein secG, similar to many e.g.  
P39388[SECG\_MYCLE PROBABLE PROTEIN-EXPORT MEMBRANE (77  
aa), FASTA scores: opt: 450, E(): 6.7e-24, (96.1% identity  
in 77 aa overlap). Start changed since original submission  
(-40 aa). PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION  
APPARATUS WHICH COMPRISE SECA[Rv3240C, SECD[Rv2587c,  
SECE[Rv0638, SECF[Rv2586c, SECG AND SECY[Rv0732."  
/codon\_start=1  
/transl\_table=11  
/product="Probable protein-export membrane protein secG"  
/protein\_id="CAD96142.1"  
/db\_xref="GI:31618224"

gene

STVVERNLRLTLFVTVGINLVSLIGVALLKYR  
complement (537. .2012)  
/gene="PE\_PGRS26"  
/locus\_tag="Mb1476c"  
complement (537. .2012)  
/gene="PE\_PGRS26"  
/locus\_tag="Mb1476c"  
note="Mb1476c, PE\_PGRS26, len: 491 aa. Equivalent to  
Rv1441c, len: 491 aa, from Mycobacterium tuberculosis  
strain H37Rv, (99.6% identity in 491 aa overlap). Member  
of the Mycobacterium tuberculosis PE family, PGRS  
subfamily of gly-rich proteins, similar to  
YODP MYCTU[O50615 hypothetical glycine-rich 40.8 kd  
protein (498 aa), fasta scores: opt: 1625, E(): 0, (55.2%  
identity in 518 aa overlap)."  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="PE-PGRS FAMILY PROTEIN"  
/protein\_id="CAD96143.1"  
/db\_xref="GI:31618225"

CDS

2119. 4419  
/gene="bisc"  
/locus\_tag="Mb1477"

gene

2119. 4419  
/gene="bisc"  
/locus\_tag="Mb1477"

CDS

2119. 4419  
/gene="bisc"  
/locus\_tag="Mb1477"

gene

2119. 4419  
/gene="bisc"  
/locus\_tag="Mb1477"

```

CDS
2119..4419
/gene="bisC"
/locus_tag="Mb1477"
/EC_number="1.-.-.-"
/notes="Mb1477, bisC, len: 766 aa. Equivalent to Rv1442,
len: 766 aa, from Mycobacterium tuberculosis strain H37Rv,
(100% identity in 766 aa overlap). Probable bisC, Biotin
sulfide reductase (EC 1.-.-.-), similar to
BisC EC01|p20099 biotin sulfoxide reductase from
Escherichia coli (739 aa), FASTA scores: opt: 1271, E(): 0,
(40.2% identity in 744 aa overlap)."
/codon_start=1
/transl_table=11
/product="PROBABLE BIOTIN SULFOXIDE REDUCTASE BIS C (BDS
reductase) (BISO reductase)"
/protein_id="CAD96144.1"
/db_xref="GI:31618226"
/translation="MQVYTSATHGVFTARVHGDDIAVAALASDTPNPAPQLQNLPGA
VHRGRIANPVRVRLQHPGSSARGAEFEVSVDELIELLASELRITVDYRGNF
AIYGSYGNASAGRFHQAQVHRFLNMLGGVYTSRHSYSAGASEVIFPHIVGAALFE
ALAEITTDVIVDHTALLVAGGLPVKNTAVMGGTTAHPDRDYVGRYARGRLVSV
SPRLDDIAALGDCRCHWLAAPVCTDVAIMGLAVLATSLADRAFLGRVCTGYE
RFRVLLGLDDGIPKTPPEAALSGLAAGDLRLARRMAHRFLITISLSLQRIERGE
QTVWAAIILAMGQIGLPGGFGHGYSSNGVNPPLACGLPALPGQNNPVSFTIPVA
AISELQPGQRLAYNGRLLELPDIKCVYAGNPNPHHQNLPRLRRALSVDYTIYVH
EYWTAMAKHADVIVTTTSFERDDPAASKNTPLTAMPVNPYANARDYHTFSAL
AHLRIGKQFTGRSARELHMYDKWAELODPVPSFAEFTWRTGLELPTRTGLTWL
ADPRDGAHPHGLTSPGRTIEFSDTVDAFALPCAGHPWTYRSEWLGPGPRAARYPLH
LIANQRTRLSHLDHGASMASKIRGPRIRIHPDDAARELTGDIYRVENDRGAC
LAGVVIDGRLFKVQVAGPDPADPRDPSMCMVGNPNLNSDGSSTSLAHGSTG
QHVLVQIERTGELPVRAHEPPLRA"
/complement(4535..5020)
/locus_tag="Mb1478c"
/locus_tag="Mb1478c"
/locus_tag="Mb1478c"
/notes="Mb1478c, -, len: 161 aa. Equivalent to Rv1443c,
len: 161 aa, from Mycobacterium tuberculosis strain H37Rv,
(100% identity in 161 aa overlap). Hypothetical unknown
protein."
/codon_start=1
/transl_table=11
/product="HYPOTHEITICAL PROTEIN"
/protein_id="CAD96145.1"
/db_xref="GI:31618227"
/translations="WGVYAEPLVIRQSVAAPAOVWQVVTPEGINDELRPMTWS
VPGAGKMTVDVPGIAGRLGRLGLVPLPDYDLSIAELRGRFREDSTMLSNR
GQWQHEVTIPEGDKTIVRDTITFQTRAGLRFAAPLIAAGLRALFGHRRRLQRFAQ
G"
/complement(5634..6044)
/locus_tag="Mb1479c"
/locus_tag="Mb1479c"
/locus_tag="Mb1479c"
/notes="Mb1479c, -, len: 136 aa. Equivalent to Rv1444c,
len: 136 aa, from Mycobacterium tuberculosis strain H37Rv,
(100% identity in 136 aa overlap). Hypothetical unknown
protein."
/codon_start=1
/transl_table=11
/product="HYPOTHEITICAL PROTEIN"
/protein_id="CAD96146.1"
/db_xref="GI:31618228"
/translations="MTVMADRSRGPAPVRRRMKTLTQAALNADKTVQVBDVLDGLGK
TMALNSSLQSLSNLTVERLDGLHLEGLTHSLDLDLAKRLILVLPVEAIVDRIDYIV
SLGETVMSPLSVTEHVRGVLDRLRNRTVHEPTN"
/complement(6061..6804)
/gene="devB"
/locus_tag="Mb1480c"
/locus_tag="Mb1480c"
/locus_tag="Mb1480c"
/notes="devB"
/complement(6061..6804)
/gene="devB"
/locus_tag="Mb1480c"
/locus_tag="Mb1480c"
/EC_number="3.1.1.31"
/notes="Mb1480c, devB, len: 247 aa. Equivalent to Rv1445c,
len: 247 aa, from Mycobacterium tuberculosis strain H37Rv,
(100% identity in 247 aa overlap). Possible devB (PGL),
6-phosphogluconolactonase (EC 3.1.1.31), belongs to a
different family to the upstream gene zwf2. Similar to
e.g. DEVB_ANASP|P46016 putative glucose-6-phosphate
1-dehydrogenase (239 aa), FASTA scores: opt: 439, E():
2.6e-20, (34.0% identity in 247 aa overlap). BELONGS TO
THE GLUCOSAMINE/GALACTOSAMINE-6-PHOSPHATE ISOMERASE
FAMILY. 6-PHOSPHOGLUCONOLACTONASE SUBFAMILY."
/codon_start=1
/transl_table=11
/product="PROBABLE 6-PHOSPHOGLUCONOLACTONASE DEVB (6PGL)"
/protein_id="CAD96147.1"
/db_xref="GI:31618229"
/translation="MSSSIEIFPDSILVAAGKRLVGAIGAAVARGQALIVLTGGG
NGIALRLYLSAQAOQIEWSKVHLFGWDERVYPEDDNLKQARRALLNHVDIPSNQV
HPMAASDGGDGLDAAALAYQVLAASAAPGDPAPNFDVHLGMPGKHNSLFPHS
PAVLSTRVWVAVDSPKPPPRITITLPAIQSRREVLLVSGFKADAVAAALGAD
PVSVPAAVAVGQNTLWLLDRDAAKLPS"
/complement(6801..7712)
/gene="opca"
/locus_tag="Mb1481c"
/locus_tag="Mb1481c"
/locus_tag="Mb1481c"
/notes="Mb1481c, opca, len: 303 aa. Equivalent to Rv1446c,
len: 303 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.7% identity in 303 aa overlap). Putative opca, OxpP
cycle protein. Highly similar to S72774 B1496.F1.30
protein from Mycobacterium leprae (265 aa), FASTA scores:
opt: 1056, E(): 0, (70.3% identity in 239 aa overlap).
Also similar to OPCA_NOSS2|P48971 putative oxpccycle
protein opca from Nostoc punctiforme (465 aa), fasta
scores: opt: 177, E(): 7.3e-05, (23.4% identity in 321 aa
Query Match 4.6%; Score 23; DB 1; Length 300050;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 193 ACCTGATCTTCGACTACCCAC 215
|||||
DB 252783 ACCTGATCTTCGACTACCCAC 252805
|||||

RESULT 18
AP005024/c 302675 bp DNA linear BCT 10-MAY-2003
LOCUS Streptomyces avermitilis genomic DNA, complete genome, section
4/30.
DEFINITION AP005024 BA000030
VERSION AP005024.1 GI:29604389
KEYWORDS Streptomyces avermitilis MA-4680
SOURCE Streptomyces avermitilis MA-4680
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
2 JOURNAL 21477403
MEDLINE 11572948
PUBMED
REFERENCE
1 Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H.,
Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.
Complete genome sequence and comparative analysis of the industrial
microorganism Streptomyces avermitilis
Nat. Biotechnol. 21 (5), 526-531 (2003)
PUBMED 12692562
```

```

3 (bases 1 to 302675)
REFERENCE
  Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
  Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
  Kushiida,N., Director-General of Biotechnology Center, Shiba,T.,
  Sakaki,Y. and Hattori,M.
TITLE
  Direct Submission
JOURNAL
  Submitted (29-MAR-2002) Director-General of Biotechnology Center,
  National Institute of Technology and Evaluation, Biotechnology
  Center, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
  (E-mail:bioenite.go.jp, URL:http://www.bio.nite.go.jp/,
  Tel:81-3-3481-1933, Fax:81-3-3481-8424)
COMMENT
  This work was done in collaboration with Haruo Ikeda(*1), Jun
  Ishikawa(*2), Akihiro Hanamoto(*3), Chigusa Takahashi(*3), Mayumi
  Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi
  Osonoe(*4), Norihiro Kushiida(*4), Hisashi Kikuchi(*4), Tadayoshi
  Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7)
  and Satoshi Omura(*1,*3).
  Final finishing process and all annotation were done by H. Ikeda
  and J. Ishikawa.
  *1 Kitasato Institute for Life Sciences, Kitasato University
  *2 National Institute of Infectious Diseases
  *3 The Kitasato Institute
  *4 National Institute of Technology and Evaluation
  *5 School of Science, Kitasato University
  *6 Institute of Medical Science, University of Tokyo
  *7 RIKEN, Genomic Sciences Center
  Following url is also available.
  http://avermitilis.ls.kitasato-u.ac.jp.
FEATURES
  source
    location/Qualifiers
    1. 302675
    /organism="Streptomyces avermitilis MA-4680"
    /mol_type="genomic DNA"
    /strain="MA-4680"
    /db_xref="taxon:227882"
    /note="This strain is also named as strain: ATCC 31267,
    NCIMB 12804 or NRRL 8165.-synonym: Streptomyces
    avermectinus"
    12. 425
    /note="SAV750"
    /codon_start=1
    /transl_table=11
    /product="putative TetR-family transcriptional regulator"
    /protein_id="BAC68460.1"
    /db_xref="GI:29604390"
    /translation="MRTVVEEELLLVDPETGPKALSTAVLARAQVDPDQVPEKE
    LRGOMAEATPOTDMAALGIVRCRKEARHAGEAGCAVAALATSLPVPSPATAMN
    RYOMARQYGIAMOEQLTCGCHVAVESDEGVAVVDRIPLPVLVALSANSFPW
    QCRDSSYSRSRWGRNPSAGPTFLFSGPERYHRQVADWATGVLLDDGMVYDFARL
    SHRYPTVIRVADVCLEHPTAQILATLARGLVESAARDWRAGREPEDSVSLRLAAG
    QARAGLSGELLHPVTMRLRASVVRALLHVGDALADAGDLIAHEAVAELELERN
    GARVORELLERTGSDRTVTECVRQTQG"
    complement(3011..3400)
    /note="SAV752"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAC68461.1"
    /db_xref="GI:29604391"
    /translation="MRTVVEEELLLVDPETGPKALSTAVLARAQVDPDQVPEKE
    LRGOMAEATPOTDMAALGIVRCRKEARHAGEAGCAVAALATSLPVPSPATAMN
    RYOMARQYGIAMOEQLTCGCHVAVESDEGVAVVDRIPLPVLVALSANSFPW
    QCRDSSYSRSRWGRNPSAGPTFLFSGPERYHRQVADWATGVLLDDGMVYDFARL
    SHRYPTVIRVADVCLEHPTAQILATLARGLVESAARDWRAGREPEDSVSLRLAAG
    QARAGLSGELLHPVTMRLRASVVRALLHVGDALADAGDLIAHEAVAELELERN
    GARVORELLERTGSDRTVTECVRQTQG"
    complement(3011..3400)
    /note="SAV753"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAC68462.1"
    /db_xref="GI:29604392"
    /translation="MQITASTVSLTVADVAASQAFILTAHLGYTVQAAADGFASLTRAD
    AVDVLARGTVLPPEQRDRAGLLIAFLTADGLHQEKLREAGVEITWPLREEP
    WGERLFIQITDPNGVIIQFVWEAYSDDT"
    complement(3457..3648)
    /note="SAV754"
    /codon_start=1
    /transl_table=11
    /product="putative TetR-family transcriptional regulator"
    /protein_id="BAC68463.1"
    /db_xref="GI:29604393"
    /translation="MFEIDRTTVSALTIGLGEVPAGRDVDPAPRGYDLRLAADRA
    VGRGRFRVVRNIIPRRRGY"
    3755..4408
    /note="SAV755"
    /codon_start=1
    /transl_table=11
    /product="putative TetR-family transcriptional regulator"
    /protein_id="BAC68464.1"
    /db_xref="GI:29604394"
    /translation="MGTEITSPQPVGLRSEKHEHETQLISDHATRLFLAQGFQTTI
    AETAAARMAKTVTYNFARKEDALDHQEEFTASLANAVADRRPGSALAALRAFH
    DSLATQNPVAGFSGPAFARMVADSPITACLRGLHDQREHLAQALADATDTTDDIT
    VHTAAGLLGTIVHRVLFQRIQDLTLAGHPNDRIHTITREATQAFDLLEPALAHYATA"
    4587..5627
    /note="SAV756"
    /codon_start=1
    /transl_table=11
    /product="putative dehydrogenase"
    /protein_id="BAC68465.1"
    /db_xref="GI:29604395"
    /translation="MTSPKTRTTTSPVACPATGSPSSQATSPAPYPAVTRACRSTWT
    CRGRDIRPAAPVPSQADFEADFPALVSRSLTDLTCGERHVLRSVSLDSFDVP
    LVTFACGTAHDAPALRLTERPFVRENPGTCRGAARPAPPGAHPRMWSVGRPASDGP
    APTSRRAASAPAGPNQHRGSANVTGTEVTPKEWLAVALVLDVHHTGTGAALRHLS
    AARGSNVGSAGGIDGDIYGAIAVNAAGAVNLTFRMALDHAPEFVRVNAHGPFTHLSG
    ITAGATPTTVRSFYAERIPLGRVAEPDEIASVVAFVLAGDPDALDYRNSNPGRRPHRC
    YRAAPHRDVNAS"
    complement(6101..8071)
    /note="SAV757"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAC68466.1"
    /db_xref="GI:29604396"
    /translation="MERSKDRPVSLQDRARYFDFRLARSTGLMGVLVILCLAVV
    PVSALLVWTDGSPRSLSGLAAVWRSASFTLRGLGTGTGLRMLLSALLGLVALLCV
    STLGVITTLGLAERLBSRGSTVLSQGHAVGLGSDQVSTVVGELVNAOSSYRPPRA
    VVLAERDKTEMERALAAHVGPAGTRTLCVSPASDPGLVSLVSPQASTVILPSPG
    EPTADRVLLALRAVLGEGTGPVLAALDDRYAPARLARGPTVLETDTVT
    ARLIAOCVGRPLSLRLDLDFAGDEFLAEATAFHGPGFCAALLGHATSCVGLLT
    AEGSTLNPAAATIVAPGSLVLTTRDGSAPEDCRHLVPEPSAIAAOPPEDAHL
    LLLGNRRAPLVNQLRTARPAGSVLDTVDVTRAPVGPTRPEPSGAARLAVRFQADLS
    RPETLLGLDLAPYDGLVILGPEPGDPRDRLTLLALLLEERTGTRMRYTEL
    TDDRRLPAPVNPVGSIVVSGELTGLLLAQIAQNRHLAPVDFELFSADGSTCLRQAG
    HYVTRGCTTPTATVVAARDRGECAIGYRRHRRRAIPDPHGIRLNPDKSRERWSADD
    QVVVATEPTPAPAAGQDTBEHQGPADTRA"
    9086..10072
    /note="SAV758"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAC68467.1"
    /db_xref="GI:29604397"
    /translation="MGRKGGIGALITVSLVGLASACGGSGDKANGDDAKASCKPG
    ASASPKPKPAGPMLLETITPTQTGTVGVAMPISVVTNPAKARAARAEKHKVSA
    SQPVSGAVHWGDKRADWRPKTYWPSGKVKIDAMKGVNNGNGRIGVHSYHTFKIG
    DDVADRVSVGTGTMKTVTRKVRVRLTSINAGSAQYPTWNGMTAVIDKQKRVHTSCSV
    GISCEKSPNYDLTLPMDWHLTQSTGYTVSTGDPTEPGSARGSGHGCVHLSKSDAK
    WFGYQVKQGDPTVITGSPRAKAPADNGYADFNLGWDQWLAGSASGEKTTATL"
    complement(10889..11494)
    /note="SAV759"
    /codon_start=1
    /transl_table=11
    /product="putative acetyltransferase"
    /protein_id="BAC68468.1"
    /db_xref="GI:29604398"
    /translation="MQRAVPRALIMAVMARPQIIECGDFVLRSGRQDDPAPAFKLIE
    ESLDHLREPWPWFARHSEESTRDLFLAKSESKWANGDVVNYAIAKDGTLIGMCQTYRGA
  
```

EPOGRRIGWHLHPAATGOGIATRAATAAMVTEMFTLPDVEYLEVAHDLANTSSAIPRR  
 LGFTVREQVTEPPAAGSGGIDVVWRLNRPTPPVDGSRP  
 11580..12161  
 /note="SAV759"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC68469.1"  
 /db\_xref="GI:29604399"  
 /translation="MRSRAPTRCWPGSTRPSDATALCWRWRWRTGCTGTGPRRFDRI  
 SAACFFRAVPPALLQAPGGKILLTGLWLYGARVLLTVADDDTAGBQKRGDTVSF  
 MSARHAPTYGNPAHWATGPEKATRTTRHSPEKITATDEALHLRFLAQSAVSDAQM  
 TTGVGVHLVDVVTGSRDHPWRGQLGCTGRG"  
 complement(12097..12768)  
 /note="SAV760"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative multicopper oxidase"  
 /protein\_id="BAC68470.1"  
 /db\_xref="GI:29604400"  
 /translation="WQAFSPGCFHAEVDYATLPSRAKNEALYGVNDRSCMLAVHD  
 HMGSMALNVAGLGLHVRDPDERLGLPGAEFVLLIRDRTFNQDGLAYNTA  
 REGDTPVNGKAYFLAVPEPRYLRLINASNEPFWRLRFDVPRDVLQGPFWLI  
 GTDGGFRAPLKMFLISAERYDLIVDFSGMPRTSRCHSFTGLAYSPTVLARGE  
 GRGSG"  
 complement(12621..12968)  
 CDS  
 Query Match 4.6%; Score 23; DB 1; Length 302675;  
 Best Local Similarity 100.0%; Pred. No. 3.9;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 193 ACGCTGATCTCGACTACCCAC 215  
 |||||  
 DB 297929 ACGCTGATCTCGACTACCCAC 297907  
 |||||  
 RESULT 19  
 BX842577  
 LOCUS BX842577 347496 bp DNA circular BCT 21-NOV-2003  
 Mycobacterium tuberculosis H37Rv complete genome; segment 6/13.  
 ACCESSION BX842577 AL022000 AL022001 AL022002 AL022003 AL022021 AL123456  
 274020 278020 281360 285982 295554 295586 295617 295890 296073  
 297345 298268 298943  
 BX842577.1 GI:38684030  
 complete genome.  
 SOURCE Mycobacterium tuberculosis H37Rv  
 ORGANISM Mycobacterium tuberculosis H37Rv  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.  
 REFERENCE 1  
 Cole, S.T., Broesch, R., Parkhill, J., Garnier, T., Churcher, C.,  
 Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,  
 Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
 Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,  
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,  
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and  
 Barrell, B.G.  
 Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence  
 NATURE 393 (6695), 537-544 (1998)  
 JOURNAL MEDLINE 98295987  
 PUBMED 9634230  
 REFERENCE 2  
 Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.  
 Re-annotation of the genome sequence of Mycobacterium tuberculosis  
 H37Rv  
 MICROBIOLOGY (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)  
 JOURNAL MEDLINE 22255591  
 PUBMED 12368430  
 REFERENCE 3  
 (bases 1 to 347496)

Parkhill, J.  
 Direct Submission  
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium  
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome  
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique  
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
 On or before Dec 4, 2003 this sequence version replaced gi:3261541,  
 gi:3261543, gi:3261544, gi:3261547, gi:3250699, gi:3261584,  
 gi:3261625, gi:3261654, gi:3261718, gi:3261771, gi:3261785,  
 gi:3242249, gi:3242245, gi:3242257, gi:3261824, gi:3261839,  
 gi:3261842.  
 Notes:  
 Details of M. tuberculosis sequencing at the Sanger Centre are  
 available on the World Wide Web.  
 (URL, http://www.sanger.ac.uk/Projects/M\_tuberculosis/).  
 FEATURES  
 source  
 1..347496  
 /organism="Mycobacterium tuberculosis H37Rv"  
 /mol\_type="genomic DNA"  
 /strain="H37Rv"  
 /db\_xref="taxon:83332"  
 116..1219  
 /gene="adh"  
 /locus\_tag="Rv1530"  
 116..1219  
 /gene="adh"  
 /locus\_tag="Rv1530"  
 /EC\_number="1.1.1.1"  
 /function="Catalyzes the reversible oxidation of ethanol  
 to acetaldehyde with the concomitant reduction of NAD"  
 /note="Rv1530, (MTV045.04), len: 367 aa. Probable adh,  
 alcohol dehydrogenase (EC 1.1.1.1), zinc-dependent,  
 similar to many e.g. AE0009|AE000958|23 Archaeoglobus  
 fulgidus section 1 (402 aa), FASTA scores: Opt: 423, E():  
 1.8e-19, (31.7% identity in 341 aa overlap). Contains  
 PS00059 Zinc-containing alcohol dehydrogenases signature.  
 TBPase score is 0.919."  
 /codon\_start=1  
 /transl\_table=11  
 /product="Probable alcohol dehydrogenase adh"  
 /protein\_id="CAAL7590.1"  
 /db\_xref="GI:2916953"  
 /translation="MSDGAIVRLVLEAPRLVRQYLRPLRGDDALVRVACGLCG  
 TDHGYTGLAGFAFVPGHETVGTAAIGPRAEORWGVSGADRVAVEVFOSCROCAN  
 CRGGRVRCVRHGLADMVGFIVDREPGLWGYAEYQYLAPDSWMLRVAGDLSPEVAT  
 LFNPIGAGIRWGVITPETKPGDVVAIVGIRGLCAAAAGAGAGFVMTGLGPRDA  
 DLALAAQFGADLVADVADDDVAALDTQTGLADVVDVTAKAPAFQAALALAPPA  
 GTVVAGTRGVSGAPGFSFEDVYVFKELRVLGALGVDTAYRAALDLVLSGRYPFASL  
 PRRCVRLGAEDLLATMAGERDGVPPIHGVLP"  
 302..346  
 /gene="adh"  
 /locus\_tag="Rv1530"  
 /note="PS00059 Zinc-containing alcohol dehydrogenases  
 signature"  
 1216..1782  
 /locus\_tag="Rv1531"  
 1216..1782  
 /locus\_tag="Rv1531"  
 /function="UNKNOWN"  
 /note="Rv1531, (MTV045.05), len: 188 aa. Conserved  
 hypothetical protein, similar to Rv0464c|MTV038.08c (190  
 aa), FASTA scores: E(): 4.8e-10, (30.9% identity in 175 aa  
 overlap)."  
 /codon\_start=1  
 /transl\_table=11  
 /product="CONSERVED HYPOTHETICAL PROTEIN"  
 /protein\_id="CAAL7591.1"  
 /db\_xref="GI:2916954"  
 /translation="MTTSRVPLLPVDEAKAAADAGVDPYMAELSTIFQVLLNHPRLAR  
 TFNDLLATMLWHGTLDLSRLREIVIMRIGWLITCDYEWQHWRAVSGLSGASDILLGVR  
 DWQGVNGFGPAEQAVLAATDDVVREGAVSAQSWACERELHCDKVVLLIELVTVISAR  
 MVASILHSLEVPEDGVSSWPPDGLSPR"  
 misc\_feature







Kodama, T., Kojima, K., Kuroaki, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kuroaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Narikawa, R., Niikura, J., Nishi, K., Nomura, M., Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, M., Numaoki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otomo, Y., Ry, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)

**COMMENT**  
This clone is one of the 28K full-length cDNA clones from japonica rice.  
URL : <http://cdna01.dna.affrc.go.jp/cDNA/>  
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.

**FEATURES**  
Location/Qualifiers  
1..723  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="001-117-C12"

**ORIGIN**  
Query Match 4.4%; Score 22; DB 8; Length 723;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 62 CGGTGCTCGGCCACAGCGCGC 83  
|||||

**Db** 71 CGGTGCTCGGCCACAGCGCGC 92  
|||||

**RESULT 21**  
AK107232/c  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) cDNA clone:002-125-B05, full insert sequence.  
ACCESSION  
AK107232  
VERSION  
AK107232.1 GI:32992441

**KEYWORDS**  
**SOURCE**  
**ORGANISM**

FLI CDNA; oligo capping.  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Erihartoideae; Oryzeae; Oryza.

**REFERENCE**  
**AUTHORS**

The Rice Full-length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-length cDNA Project Team:  
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group; Otomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ry, R., Ueda, M., Matsubara, K., RIKEN:  
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M. and Hayashizaki, Y.

**TITLE**  
**JOURNAL**

Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice  
Science 301 (5631), 376-379 (2003)  
22752273  
MEDLINE  
12869764

**REFERENCE**  
**AUTHORS**

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kuroaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Negata, T., Nakamura, M., Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K., Numaoki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otomo, Y., Ry, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

**TITLE**  
**JOURNAL**

Direct Submission  
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)

**COMMENT**

This clone is one of the 28K full-length cDNA clones from japonica rice.  
URL : <http://cdna01.dna.affrc.go.jp/cDNA/>  
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Otomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ry, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.  
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,

```

Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Iida, J., Inamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A. and Hayashizaki, Y.

FEATURES             source
1. 1019
   Location/Qualifiers
   /organism="Oryza sativa (japonica cultivar-group)"
   /mol_type="mRNA"
   /cultivar="Nipponbare"
   /db_xref="taxon:39947"
   /clone="002-125-E05"

ORIGIN
Query Match          4.4%; Score 22; DB 8; Length 1019;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 62 CGGTGCTCGGCCACAGCGCGC 83
    |||||
Db 589 CGGTGCTCGGCCACAGCGCGC 568

RESULT 22
LOCUS AC135598 137312 bp DNA linear HTG 16-OCT-2003
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 3 clone
ACCESSION OSJNB0021K20, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.
VERSION AC135598
KEYWORDS HTG; HTGS PHASE2; HTGS ACTIVEFIN.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 137312)
AUTHORS Buell, R., Hsiao, J., Zismann, V., Moffat, K.M., Hill, J.,
Gausberger, K., Burgess, S., Jarrahi, B., Shvartsbeyn, M., Brenner, M.,
Ciecko, A., Pal, G., Vanaken, S., Hansen, C., Utterbach, T.,
Feildlyum, T., Khalak, H.G., Yuan, Q., Quackenbush, J., White, O.,
Salzberg, S. and Fraser, C.
TITLE Oryza sativa ssp. japonica cv. Nipponbare OSJNB0021K20 BAC genomic
sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 137312)
AUTHORS Buell, R.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 137312)
AUTHORS Buell, R.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
COMMENT On Oct 16, 2003 this sequence version replaced gi:24418097.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 68692: contig of 68692 bp in length

```

```

* 68693 68792: gap of unknown length
* 68793 137312: contig of 68520 bp in length.

FEATURES             source
1. 137312
   Location/Qualifiers
   /organism="Oryza sativa (japonica cultivar-group)"
   /mol_type="genomic DNA"
   /cultivar="Nipponbare"
   /sub_species="japonica"
   /db_xref="taxon:39947"
   /chromosome="3"
   /clone="OSJNB0021K20"

ORIGIN
Query Match          4.4%; Score 22; DB 2; Length 137312;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 62 CGGTGCTCGGCCACAGCGCGC 83
    |||||
Db 111168 CGGTGCTCGGCCACAGCGCGC 111189

RESULT 23
LOCUS BX572597 349737 bp DNA linear BCT 18-DEC-2003
DEFINITION Rhodopseudomonas palustris CGA009 complete genome; segment 5/16.
ACCESSION BX572597 BX571963
VERSION BX572597.1 GI:39648199
KEYWORDS complete genomes.
SOURCE Rhodopseudomonas palustris CGA009
ORGANISM Rhodopseudomonas palustris CGA009
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiales; Rhodopseudomonas.
REFERENCE 1 (bases 1 to 349737)
AUTHORS Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L.,
Land, M.L., Pelletier, D.A., Beatty, J.J., Lang, A.S., Tabita, F.R.,
Gibson, J.L., Hanson, T.E., Torres y Torres, J., Peres, C.,
Harrison, F.H., Gibson, J. and Harwood, C.S.
TITLE Complete genome sequence of the metabolically versatile
photosynthetic bacterium Rhodopseudomonas palustris
Unpublished
REFERENCE 2 (bases 1 to 349737)
AUTHORS Larimer, F.W. and Harwood, C.S.
CONSRTM Rhodopseudomonas genome consortium
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2003) Submitted on behalf of the Rhodopseudomonas
genome consortium, the DOE Joint Genome Institute, Production
Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
larimerf@ornl.gov

FEATURES             source
1. 349737
   Location/Qualifiers
   /organism="Rhodopseudomonas palustris CGA009"
   /mol_type="genomic DNA"
   /strains="CGA009"
   /db_xref="taxon:258594"
   complement(232..1425)
   /gene="carA"
   /locus_tag="RPA1276"
   complement(232..1425)
   /gene="carA"
   /locus_tag="RPA1276"
   /EC_number="6.3.5"
   /function="InterPro
IPR000991:IPR001317:IPR001687:IPR002474:IPR006274 COGS
COG0505"
   /note="observed by proteomics
Citation: Proteomics from VerBerkmoes et al. (2003)
unpublished"
   /codon_start=1
   /evidence=not_experimental
   /transl_table=11

```

```

/product="carbamoyl-phosphate synthase small subunit"
/protein_id="CAE26719.1"
/db_xref="GI:39648200"
/translation="MTNSASHPAWPDHKKHTALLVLADGVTFVFGFGLGARGHAVGEVCF
NTMTGYSIELTDPSTAGQLIITFTPHIGNVGTNDIEDTNNMAATPGARGVILRSAL
TDSNYRSHLDLWAKGIIIGLSGIDTRALTALIRSGMPNAVIAHSPTGEFDLHA
LKEAREWPLGEMDLVPWGTQAFRTWDETFWMAGESFGRODKPEFNVAIDYGIKR
NIRLLAGGCKVTVVPATTSADILAMKPDGVFLSNGPDPAATCKYAVPVHGVIE
TGVPFGICLGHQMLGAGKTVKMHQGHGANHPKDLTTGKVEITSMHGFAYDK
SLTPANVQOOTHLSLFDSDNCIALADKPVFSVOYHPFASPGQDSHYLFRFTDLNRA
KKAAS"
gene 1574..2278
/locus_tag="RPA1277"
CDS 1574..2278
/locus_tag="RPA1277"
/function="InterPro
IPR000073:IPR000379:IPR001031:IPR003089 COGs COG0596"
/notes="observed by proteomics
Citation: Proteomics from VerBerkmoes et al. (2003)
unpublished"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="possible beta-ketoadipate enol-lactone
hydrolase"
/protein_id="CAE26720.1"
/db_xref="GI:39648201"
/translation="MTSLIPVLVPGIAGSPRIYAEIFFALMACGSPVTVANHIRDS
MAAIRILAEAPRFALAGSHGGYIAEIMQAPDRVAKLALINTQALPDSPAEATE
RRTQIGEAAGRLHVDLDEYGFVHPHRRDDAVLRTVVHEMGDDVGAAGFVQORA
IARTDSRHTGLTGCTVLIVISGDTTTPINSLSQEMADGITGAKLLIIPDCGHLPOI
EOPAAVAALADWLNR"
gene 2436..2891
/locus_tag="RPA1278"
CDS 2436..2891
/locus_tag="RPA1278"
/function="InterPro IPR003789 COGs COG1610"
/notes="observed by proteomics
Citation: Proteomics from VerBerkmoes et al. (2003)
unpublished"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="GatB/Ygey"
/protein_id="CAE26721.1"
/db_xref="GI:39648202"
/translation="MURDDINNAVKEAMKAKDERKLSLRMNSTINKADIARGQGK
PLSDGDLISLQKMIKQDSVELYDKGRAELAQDRAETIAVIQALPQQMSDDEV
KAAIAATISSETGAAGIKDMGKVTGALKAKYAGQMDFGKAGMVKAAALTG"
gene 3279..3287
/locus_tag="RPA1279"
CDS 3279..3287
/locus_tag="RPA1279"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE26722.1"
/db_xref="GI:39648203"
/translation="MDAFAPQHUHVIVVSAGWOAPLHTVVNGCAGSAPRSLVGLSRL
RRAPDRRGHDPAALPPQFQAGDGLSLTSHSGFASDRDPGLIQFSRKAQRGARRP
"
gene 3478..4326
/locus_tag="RPA1280"
CDS 3478..4326
/locus_tag="RPA1280"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE26723.1"
/db_xref="GI:39648204"
/translation="MSISGKVDPVWRPTITIGDIAEALQGLRDFQAPLYGLAFAPY

```

```

AAGGLILACLTAFAHMYLALYPLGAGFALIGFVAGLYEVSQRQREAGKRPSSLIQIVG
LMRSSELGMMAFVTLFLVIMYOVRLLIALFLVGAGSGSQEETISAVLTNREGIV
FIAGNCVGCACIALVLSLTAVSFPLLRDQDVFVTAMTSVRAVVKSPMLPTIGWAAT
IVLLAISALPFLGLVVLTVLPGHATWHLVYKRVAPVAELPDTADTEASNNVAMP
KKAATG"
misc_feature 3648..3658..3726,3805..3858,3868..3936,
order (4023,4132..4200)
/locus_tag="RPA1280"
/notes="6 probable transmembrane helices predicted
byTMHMM2.0"
gene complement (4440..4913)
/locus_tag="RPA1281"
CDS complement (4440..4913)
/locus_tag="RPA1281"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE26724.1"
/db_xref="GI:39648205"
/translation="MSKPRFAVYVYRTTPDLRHLALTDADETRRYMDCTLRSDMTV
GSEMTMERGEIKNRCTIYESDPHRLAYDWSVWDPAMQEKPSVTVYIEPQGLV
KUTVHQNTQGSITLPSISFGFMVLSKLSLETGQLPFPAPQSHAAESAHV"
gene complement (4906..5220)
/locus_tag="RPA1282"
CDS complement (4906..5220)
/locus_tag="RPA1282"
/function="InterPro IPR001845 COGs COG0640"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="possible transcriptional regulator, ArsR family"
/protein_id="CAE26725.1"
/db_xref="GI:39648206"
/translation="MDEVFALADGSRRLDLRHARGQTLSDLCGLAMTQAVTK
HLAILEAANLVATVKHREKLYLNPMPHEIGERWIRKFERGLAALSDLKRLKETR
DE"
gene 5349..5966
/locus_tag="RPA1283"
CDS 5349..5966
/locus_tag="RPA1283"
/function="InterPro IPR001123 COGs COG1280"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="homoserine/homoserine lactone/threonine efflux
protein"
/protein_id="CAE26726.1"
/db_xref="GI:39648207"
/translation="MSLEMYLTFVTACIVIALIPGPVTVIANGLRHGTAAALNIA
GAQLGGLVIALVAAGLTTLMTMGVFDWVFLGAAYLVMLGILRRVEGLSTDQ
PPTPRGGPFLQGFVLLANPKVLFVFGAFIPQFVDMQRETVPQVALLGVTFVIAAM
TDAYALLAGRAKTMFSQORILLRSLRISGFMIGGIGIWLALSAK"
sig_peptide 5349..5426
/locus_tag="RPA1283"
/notes="Signal predicted by SignalP 2.0 HMM (Signal
peptideprobability 0.832) with cleavage site probability
0.732 atresidue 26"
misc_feature 5435,5463..5531,5568..5621,5679..5747,
order (5367..5435,5463..5531,5568..5621,5679..5747,
5784..5852,5895..5954)
/locus_tag="RPA1283"
/notes="6 probable transmembrane helices predicted
byTMHMM2.0"
gene complement (6090..6887)
/locus_tag="RPA1284"
CDS complement (6090..6887)
/locus_tag="RPA1284"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved unknown protein"
/protein_id="CAE26727.1"
/db_xref="GI:39648208"

```

/translation="MDGVDRNFLSLDAGEPPVIERNONGOSPFLVCDHYGRQIPQAOL  
GDGLGPAETRIHAWDGTAGVAERMSDALGAHLIAQYSLRVIDNCNRPSPASISIP  
LVSEATIPNDGLSREHAEERRLLIFDPFHOREIAEIDROLSGOPTVLLALHSPTS  
VYHGMARPHIGTIGYQCTVIPPULLAELRRFEDVLGDNQPIAVSDGDTYITPVHGT  
ARGLITSGBIRQDITDEAGQVEWASRLTRVLSTIADTLRQQQALR"

Query Match 4.4%; Score 22; DB 1; Length 349737;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 357 GCTGCTCGGCTCGCGACACC 378

Db 135785 GCTGCTCGGCTCGCGACACC 135764

RESULT 24  
AF144053 727 bp DNA linear BCT 12-JUL-1999  
LOCUS  
DEFINITION Micromonospora inyoensis macrolide type polyketide synthase genes,  
partial cds.  
ACCESSION AF144053  
VERSION AF144053.1 GI:5442290  
KEYWORDS Micromonospora inyoensis  
SOURCE Micromonospora inyoensis  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE 1 (bases 1 to 727)  
Hyun, C.G. and Suh, J.W.

TITLE The use of PCR to isolate new macrolide type polyketides from  
actinomycetes

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 727)

Hyun, C.G. and Suh, J.W.

TITLE Direct Submission

JOURNAL Submitted (19-APR-1999) Department of Biological Science, Institute  
of Bioscience and Biotechnology, Namdong San 38-2, Yongin,

Kyunggi-Do, Korea

FEATURES Location/Qualifiers

source 1..727

/organism="Micromonospora inyoensis"

/mol\_type="genomic DNA"

/strain="ATCC27741"

/db\_xref="ATCC:27741"

/db\_xref="taxon:47866"

<1..567

/note="type I PKS"

/codon\_start=1

/translation="macrolide type polyketide synthase"

/protein\_id="AAD43313.1"

/db\_xref="GI:5442291"

/translation="ASRDLGDFSVLSLELRNLAATGIRLASGLLFNHPPTAALAGH  
LRDLVGGVPPAATRADADAEPYVAICRYPGVDRSSDLRLVAGGVDAIGEL  
PTNRGWDLEGLFDPPEPGVGKSYARHGGLFRADEFDEFFGISPREEAAMDPOQRLL  
LETSEWAFERAGIDPTACAAAPACSWG"

567..727

/note="type I PKS"

/codon\_start=1

/translation="macrolide type polyketide synthase"

/protein\_id="AAD43314.1"

/db\_xref="GI:5442292"

/translation="MSQEVGFLYTSAGSDGYRLTGTSVASGRISYVVFEGPAP  
TVDTACSSSL"

ORIGIN

Query Match 4.2%; Score 21; DB 1; Length 727;

Best Local Similarity 100.0%; Pred. No. 92;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 CGCGGCCACCGGCTCGGCT 185

|||||

Db 60 CGCGGCCACCGGCTCGGCT 80

RESULT 25

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

SGR300302 39314 bp DNA linear BCT 20-MAR-2002  
Streptomyces griseus partial ORF1, canA gene, canC gene, canF gene,  
canT gene, canRA gene and canRB gene.

AJ300302

AJ300302.1 GI:12231153

ABC-transporter; aminotransferase; can RB gene; canA gene; canC

gene; canF gene; canRA gene; canT gene; Cho-like protein;

cytochrome P-450; ferredoxine; PABA synthase; thioesterase.

Streptomyces griseus

Streptomyces griseus

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Streptomycineae; Streptomycetaceae; Streptomyces.

1 Campelo, A.B.

Thesis (2000) Department of Microbiologia, Universidad de Leon,

Leon, Spain

2 Campelo, A.B. and Gil, J.A.

The candidin gene cluster from Streptomyces griseus IMRU 3570

Microbiology (Reading, Engl.) 148 (Pt 1), 51-59 (2002)

21642576

11782498

3 (bases 1 to 39314)

Gil, J.A.

Direct Submission

Submitted (14-NOV-2000) Gil J.A., Microbiologia, Universidad de

Leon, Campus de Vegazana s/n, 24071, SPAIN

Location/Qualifiers

1..39314

/organism="Streptomyces griseus"

/mol\_type="genomic DNA"

/strain="IMRU 3570"

/db\_xref="taxon:1911"

complement (<1..1785)

/note="ORF1"

/codon\_start=1

/translation="cho-like protein"

/protein\_id="CAC22112.1"

/db\_xref="GI:12231154"

/db\_xref="SPTREMBL:O9EWC6"

/translation="MILLERTEIALVDAAALGAATEGRSSLLVLTGPIGIRSAALLORL  
SCLADDDVVRVRAHRAAMEQDFAGVVRQLFETLLGDSPEADRDMDHAFARH  
VIADDAAPPDAQIAATEAVLHGLLSLVANSADSLILLVDDLDQSDVFSRLWLY  
LAKELHGLRAVVCALRDGDPDRSHHTLVREIREAATQTLRPASLSLTHLVRHFG  
EAGDDEFVACHASVGNPLFLSILVTGFLGRPLAEHAEARLRPSQLRRELAS  
ILRTQAPVRDLAAAIILGQSDAPTLRLAGLDSIGYAGALRALGALGAAPDEP  
RFHRSVRDAESTLTWQREHDEAALYAGSPAEQVAOLMAVVRROPWAVD  
VIRAAATLARGAPDPAAGLTCRLALDSPPAGVGRGLLVELGTAEKGFPLACERH  
IAQMTLLPEPRDRAVALRISPTALGPAPLAVDLLRQAADLGPAGLGTDRDLA  
LRLRLAHLCHGEDPHELAASAHRLKEIGPEPPVDTGGERELTALLGAATLGHCPA  
SEVRLAGRVLDREPATSARTHTLPLAI"

9525..10583

/gene="canA"

9525..10583

/gene="canA"

/codon\_start=1

/translation="aminotransferase"

/product\_id="CAC22113.1"

/db\_xref="GI:12231155"

/db\_xref="GOA:O9EWC5"

/db\_xref="SPTREMBL:O9EWC5"

/translation="MAFTHPVSRPALDGRLELYVSDAVSGGWISSQGYVRRFEEAFA  
EWNGVAGVACSSGTALTLRALNIGPGDEVIVPEFTWASAWAVTGTATPVFVD  
CGDDLINIVTIEKLIIPRTAVMPVHYVGRCDMDVMDLALQYNLRVSDSAAHG  
VRPVGDIACFSIFANKIITAGEGVGCLTDDFLAEQLAHLRAMAFTRDHSFLHKKLAY  
NYRMTAMQGAVALAQTERTLEILATREIARYDAGLKDLPGLTLPMPARDVLNMYDLR

9525..10583

/gene="canA"

9525..10583

/gene="canA"

/codon\_start=1

/translation="aminotransferase"

/product\_id="CAC22113.1"

/db\_xref="GI:12231155"

/db\_xref="GOA:O9EWC5"

/db\_xref="SPTREMBL:O9EWC5"

/translation="MAFTHPVSRPALDGRLELYVSDAVSGGWISSQGYVRRFEEAFA  
EWNGVAGVACSSGTALTLRALNIGPGDEVIVPEFTWASAWAVTGTATPVFVD  
CGDDLINIVTIEKLIIPRTAVMPVHYVGRCDMDVMDLALQYNLRVSDSAAHG  
VRPVGDIACFSIFANKIITAGEGVGCLTDDFLAEQLAHLRAMAFTRDHSFLHKKLAY  
NYRMTAMQGAVALAQTERTLEILATREIARYDAGLKDLPGLTLPMPARDVLNMYDLR

```

AEREELRAHLDERGIERLFFKPMRQPGYLDVWPTLNAHRFSEBGLYLPTHTGLT
AAQEYITGAVRDFYRAG"
10678. .11859
/gene="canC"
10678. .11859
/gene="canC"
/codon_start=1
/transl_table=11
/product="cytochrome P-450"
/protein_id="CAC22114.1"
/db_xref="GI:12231156"
/db_xref="GOA:Q9EWC4"
/db_xref="SPTREMBL:Q9EWC4"
/transl_table="MTSPGPTVVDFFPRTPREPLPLSOYAEHRKQNGLVOTHLPNGR
PIWLTHREDYRAVLTHPRISANPDNEGPNVGTMGVQKQEQIPFGFVGLDSPEHR
FRKVLPEFTVRRVRELPALERTVDERIDAMLAGGNADLVNDFKVPVSLVISALL
GVPSADRDFFSRTVAISTDEERAEATQRLRYINRLIVIKKRGEDLISRL
LSTGKLDEELSGVLLLIAGHTTANNIGLVVTLSHREWIGDDRLVELLARLHS
VADMVALRVAVDDVEIAGQTRKGGGIVPLLASANHDTEAFGCPHAENPETERRHA
FGYGHQCLGNLVRVMEIAYKLFERIPELRLAVPEDQLAYKXDGLFGLHELPIYR
W"
11906. .12100
/gene="canF"
11906. .12100
/gene="canF"
/codon_start=1
/transl_table=11
/product="ferredoxine"
/protein_id="CAC22115.1"
/db_xref="GI:12231157"
/db_xref="GOA:Q9EWC3"
/db_xref="SPTREMBL:Q9EWC3"
/transl_table="MRVTVDEQCQVAGQCQVNAPEVFDQDDGGVVLLRAEPDERDH
EAVRTAGLCPASVVLQED"
12223. .12993
/gene="canT"
12223. .12993
/gene="canT"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="CAC22116.1"
/db_xref="GI:12231158"
/db_xref="GOA:Q9EWC2"
/db_xref="SPTREMBL:Q9EWC2"
/transl_table="MTTADDTAARLWRYHPAEADAIVRLVCFPHAGGSASFVHPVSR
FAPGAEVVSQYVPGQRKKEPCVPLDGLTLADLITEQLPLDERPTVFFGHSMAALA
FETAWRLQKAGPRTVIASGRGPGSTTAEARVHTRDDGIVAEMKRLNGTAAGVLGD
BEILMVALPALRGDVRATETVTCPPDRRLACGLVLTGEBDDPLTTVEEAERWRDHTTG
PRLRVFGGHPFLTQHLDVANTETIAQALHPPRAAPA"
13190. .15361
/gene="pabAB"
13190. .15361
/gene="pabAB"
/codon_start=1
/transl_table=11
/product="PABA synthase"
/protein_id="CAC22117.1"
/db_xref="GI:12231159"
/db_xref="GOA:Q9EWC1"
/db_xref="SPTREMBL:Q9EWC1"
/transl_table="MRTLLVDNYDSFTYNLPHYLFRANGREPEVINDDPAPWRGLLD
AFDNLVSPGPTPHRPAFGLCARIAEEGRPLVGLGVLGHQGMALHAGRVRAPEP
RGRGTSANVHDTGLFEGLPQPLEVVRVHSLATPELPELEATAWSDEQVWALRHRT
LPLGWGVPESIGTQDGHRLANFRDLTERGTRRHGGAGHGTLPPAPARETKAT
TGTPLRLVIAKSLPDAEVAFDSLPTGDPHFWLDSRRPGELGOLSMMDGASGP
LARTAKDVHACTVTRVADGASTVESAFELTWLENDLACLTVEPELPAFALGWVC
LGYELKACDGGAAHRSPPDAVLPADRALVLDHRTTYLLALVEDDAEABEAWL
AAASATLAVAGREPEPCPEAPVCTTGPVELRHDRDGYLKLIDVCOQEAAGAEYEV
LTNMAEDTFTFPFAALRVSPAPFAAFDLPFGMAVLSSEPERFLRDHGRMES
LQKIGTRPGATQPDADLALALACEKRAENIMVLDVLRHDLGRCAEVSQWADPV
FOVETVATVHVQVSTVATRLRDESSPVAARAAFPGSGMTGAPKRTMOIIDLPGGP
RGVYSSAIGYFSLTGAVDLSIVIRTVVLGSGRLRYGVGGAVIALSDPDAFEETAVKA
```

```

APLLRLDITAPPGREAPFKDLDGPDGDTAGAPKDLVLPQ"
20829. .21836
/gene="canRA"
20829. .21836
/gene="canRA"
/codon_start=1
/transl_table=11
/product="ABC-transporter"
/protein_id="CAC22118.1"
/db_xref="GI:12231160"
/db_xref="GOA:Q9EWC0"
/db_xref="SPTREMBL:Q9EWC0"
/transl_table="MNGIHLPPALPAPGVSAAAGSGDLMIEAGVSKAYGTVHALDSV
SLAVPRCTVGLIGHGAGKTTLVLDILTALPTTSGRAVAGYDVADKPVIRRRIGL
TGQASVDYDQSLGYDNLVLIARLAGARRARVARELLELFLTEVADRASSYSGG
LRRDLADIALVSGAPEVFLVDEPTTGLDPSRRINWEIVEGIVGQTTVLLTYLEE
ADRLADIALVSGRAVVAAGTAPBELKATVGRRTVTLTLETGEDVSAARTALRGAGFAP
VDGEARTVVPIDATREIADIIRSLDQAGVEASELNFGEPTLDDVYLLTAEQAAGRTR
A"
21970. .22776
/gene="canRB"
21970. .22776
/gene="canRB"
/codon_start=1
/transl_table=11
/product="ABC-transporter"
/protein_id="CAC22119.1"
/db_xref="GI:12231161"
/db_xref="GOA:Q9EWB9"
/db_xref="SPTREMBL:Q9EWB9"
/transl_table="MPRAVAPAFGSSLSFTQIRVLTGRSLRAMVTDPGIVLFGLIQP
VVILFVLTQVFSKMGMPHPFDGVSYLDVLPVAVDVAOASOMSGVGLVEDOKNGI
VARLRSFVHGALLAARSLVGLVRSVAQVAVIMALANTVGLGSPNGGAELASGL
TLTISFSGWAPFAGAWERRAEPLQNLALVIFPFASASAYVPAVDLPWLGAVAS
VNPLTYAIDATRALDVPGLDHALPAVVICAVIAVVGVMIAAVRGFRRLP"
Query Match 4.2%; Score 21; DB 1; Length 39314;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 121 GGCTTCGACTCGCTGGCCGCG 141
|||||
Db 34672 GGCTTCGACTCGCTGGCCGCG 34652
RESULT 26
AF263245
LOCUS AF263245 47981 bp DNA linear BCT 17-SEP-2000
DEFINITION Micromonospora megalomicea subsp. nigra megalomicin biosynthetic
ACCESSION AF263245
VERSION AF263245.1 GI:10179840
KEYWORDS SOURCE
ORGANISM Micromonospora megalomicea subsp. nigra
Micromonospora megalomicea subsp. nigra
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE 1 (bases 1 to 47981)
AUTHORS Volchegursky, Y., Hu, Z., Katz, L. and McDaniel, R.
TITLE Biosynthesis of the anti-parasitic agent megalomicin:
transformation of erythromycin to megalomicin in Saccharopolyspora
erythraea
JOURNAL Mol. Microbiol. 37 (4), 752-762 (2000)
MEDLINE 20430101
PUBMED 10972798
REFERENCE 2 (bases 1 to 47981)
AUTHORS McDaniel, R. and Volchegursky, Y.
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2000) Kosan Biosciences, Inc., 3828 Bay Center
Place, Hayward, CA 94545, USA
FEATURES Location/Qualifiers
1. .47981
source
```

```
/organism="Micromonospora megalomicea subsp. nigra"
/mol_type="genomic DNA"
/strain="NRRL3275"
/sub_species="nigra"
/db_xref="taxon:136926"
/complement(<1. .144)
/gene="megI"
/complement(<1. .144)
/gene="megT"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyglucose-2,3-dehydratase"
/protein_id="AAG13906.1"
/db_xref="GI:10179841"
/translation="MGDRVNGHATPESTQATRFTRHGPPPTATDDVDHDLAHRRAE
HRLK"
928. .2061
/gene="megDVI"
928. .2061
/gene="megDVI"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 3,4-isomerase"
/protein_id="AAG13907.1"
/db_xref="GI:10179842"
/translation="MAVGDRRLRGRELQMARGLYWGFGANGDLYSMLLSGRDDDPWTW
YERLRAAGRPVAGRAAGTVVGGDRTAAEVLADPGFTGPPDAARWQVAHCPAASWA
GPREFYATEDAASVTVDADLQQRCAVLTELGSREFDLVNDFAREVFLALGTAPA
LKGVDRLRSWTSATRVCLDAQVSQQOLAVTEQALTDALDEIDAVTGGDRAVLGVV
AELANVTGNVLAVTELPALADDPETATRVTVESVTSFGVHLERRTAAASRR
VGVVDPTGGEVTVVAAANRDEFTDPDFVDVDRGDAEILLSRPGSPRTDLDLV
ATLATAALRAAAVLPRLSRSGVPIRRRSPVARGLSRCPVEL"
2072. .3382
/gene="megDI"
2072. .3382
/gene="megDI"
/codon_start=1
/transl_table=11
/product="TDP-megosamine glycosyltransferase"
/protein_id="AAG13908.1"
/db_xref="GI:10179843"
/translation="MRVFFSSMAVNSHLFGLVFLASAFQAAGHEVRVVASPALTDDVT
GAGLTVPGDDVDEWHAHAGQDIVENRTLDWDQSHTTMSWDDLGMQOTFTPT
FFALMSPSLIDGNVEFCRSDREDWIWEPLTAAPIAAVGTGTPHARMWGPDAVTR
AROSFLRLAHQVEHREDPLAEWDTLRRFGDDPHLSFDEELVLGQWTVDPLEPL
RIITGVTVGWRVYVPGNPSVPVALLREPERRVCITLGSREHIGIGOVSGEMLD
AIADIDAEFVATDDQQLVGVSVPANVTAGFVPMNVLPTCAVTHHGCTGSWLTA
AIGHVPOILSDADTEVHAQQLQDLGAGLSLPVAGMTAEHLRGAIERKVLDEPAYRLGA
ERMRDGMRTDPSQAQVVGICQDLAADRAARGRPRTAEPLPR"
3462. .4634
/gene="megY"
3462. .4634
/gene="megY"
/codon_start=1
/transl_table=11
/product="mycarose O-acyltransferase"
/protein_id="AAG13909.1"
/db_xref="GI:10179844"
/translation="MVTSTNLDYTABNLSLTCMRVFAAFLVFFTHVLSRLPNYSV
YADGLDAPWQTTGRGVGSFFLISGFLVTSARASDSVWFRRRVCKLFFNHLVTAF
AAVLIVTQAVSGEALIFNLLIHAFFALEISFGINPVSWSLACEAFYLCFPLF
LFWISGRPRLMNAVAIVFAAIWAVPVVADLLPSSPPPLTGLYEYSAI QDMFLYTFP
LRSLFTLIGLILITGRWINVGLLPVLLPVFPVVASLTPGVVAISSSMYIDPA
LVLLIASCATADLOOKTEENRYVMVLGDVSFALYVMVHTLVIVYGDLLGFSQTELA
PLGALPMIPLFLAVSLVLSWLLYRFVELPVMRWAPASRRAPATEPEQTFSRR"
4651. .5775
/gene="megDII"
4651. .5775
/gene="megDII"
/codon_start=1
/transl_table=11
/product="TDP-3-keto-6-deoxyhexose 3-aminotransaminase"
/protein_id="AAG13910.1"
```

```
/db_xref="GI:10179845"
/translation="MTTYVWSYLLEYRERADILDVQKVFASGSLILQOSVENFETE
YARYGHIAHCVGVDNGTNAVLALRSVGVGRDDEVVTVNSNTAAFTVIAIDEICARPVF
VDVREDYLMDDTLVEAAVTPRTKALVPVHLXQCVDMTALRELAARGLKLVCAQ
AHGARDGELAGTMSDAAAFSPYFKVLGAYGGGAIVTDDDEBALRELARYGMEE
VYIVTRTPGHSRLDEVQBAELRRKILRLDIDAVAGRAVQRYVDGLADLQDSHGLEL
PVVTDGNEHVFFYVYVHRRDEIILKRLDGDYDILNYSYPMFPVMTGTFAHLGVASG
SLPVTTERLAGEIFSLFMPBLSLPHLDLQDRVIEAREVITGL"
5822. .6595
/gene="megDIII"
5822. .6595
/gene="megDIII"
/codon_start=1
/transl_table=11
/product="daunosaminyl-N,N'-dimethyltransferase"
/protein_id="AAG13912.1"
/db_xref="GI:10179847"
/translation="MPNSHSTSSSTDVAPYERADYIYDFYHGRGKGYARAEADALVEVA
RKHTPOATLLDVACGTGSHVELDASPREVGVGVDLSAAMLAATAANDPGREHQDGM
RDFSLDRDPDVVTCMFSYGLVDEAEELDRVANLAGHLAPGGLTVPEWPFETFRP
GWGADLVTSDDRISMSHTVPAGLPDRTASRMTIHYTVGSPGAGIEHTEVHVMTL
EAAAAYEQAFQAGRLSCSYVGHDLFSPGLFVGVAABEGR"
6592. .7197
/gene="megDIV"
6592. .7197
/gene="megDIV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 3,5-epimerase"
/protein_id="AAG13911.1"
/db_xref="GI:10179846"
/translation="MRVEBELGEGVFTFTQTFADERGVTGTAQEDVFVAALGRPLF
PVAQVSTRSRGVGVGVHFTMPGSMKYVYVARGAMDFAVDIRPGSTFGRAEPF
ELASMSVGLYLPVGMGHLFVLEDDTTLVILMSAGYVDPKERAVHPLDDELALPIPA
DLDLWMSRDRVAPTLREARDQGLPDYACRAAAHVVRT"
7220. .8206
/gene="megDV"
7220. .8206
/gene="megDV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AAG13913.1"
/db_xref="GI:10179848"
/translation="MVLGASGLGSAVTHALADLPVRVRLVARREVYPVPSGADVDE
THVDLITPCALAEVADARVFAAQIRGTSGWRI SEDDVVAERTNVLGLLEATIAV
LSRSHAPVTVFGSNTQVGRVTAGVIGSEODHPGVYDROKHTGEQLKEATAAG
AIRATSLRLLPVFGVPAAGTADRGVVSTMRRLTQQLITMHDGTVRELLYVTD
ARAFVTAHDADALAGRHLLGTGRSNWPLGVEVQAVSRSVARHTGEDFVFPVVPFPA
HMDPSDLRSVEDPAPRTAVTGWRTVWAEAVDRTVAALAPRAAAPSFS"
complement(8228. .9220)
/gene="megDVII"
complement(8228. .9220)
/gene="megDVII"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 2,3-reductase"
/protein_id="AAG13914.1"
/db_xref="GI:10179849"
/translation="MGTTAGSARVRVGRSALHTSLRLWLTGTVNFSGRVTDALRLMD
HALRGVNCIDADIYWRLYKGTBELVGRWFAQGGREETVTLATKVGSEMSRVN
DGSLRARIYACENSILRLGVGDIHYQTHDRAAPDEWVQAAEHLVSGSKGVY
GSLNAGWHLIAAQAESAARNLLGMSHQCLYNLAVRHPBLDVLPAQAYGVGVFAWS
PLHGGLSLVGLIACTAVKSAQRAOVLPAVRPLVEAYEDYCRRLGADPAEVGLA
WVLSRPGILGAVIGPTPEQLDSALRAAELTLGEEELRELEAIFPAVADGPVP"
complement(9226. .10479)
/gene="megBV"
complement(9226. .10479)
/gene="megBV"
/codon_start=1
/transl_table=11
/product="TDP-mycarose glycosyltransferase"
/protein_id="AAG13915.1"
```

```
/db_xref="GI:10179850"
/translation="MRVLLTSTFAHRTHTFOGLVPLAWALHTAGHDVVRVASOPELTDVVV
GAGTSLPLGSDHLEFDLSPAAQVHYRTTDLDFARRGPELRSMWFLHGIIEATSRF
VFPVNDSSFVDELVEFAMDRPDLVLEWPEPTFAGAAKACGAAHARLLWGSDLTGY
FRSQDLRGQRDRPDLGGWLTVEAGFRGLDYSIEDLAVGQNSVDQLPESRLET
GLESHVTRTLFYNGSSVVPQMLRSDGVRRVCFGTGSGALGITSNPQEFRLTLATLAR
FDGIVTVTRSGLDPASVPDNLVDFVPMNILLPGCAAVIHGGAGSNATLHHGVPO
ISVAHEDCVLRQGTAEAGVFLRPDEVDADTLWQALATWVEDRSHAEAKLEKROE
ALAAPTAEVVPVLEALAHQHRADR"
complement(10483..11424)
/gene="megBIV"
CDS
complement(10483..11424)
/gene="megBIV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AAG13916.1"
/db_xref="GI:10179851"
/translation="MTRHVTLLGVSGFVGSALLREFTHPLRLRAVARTGSRDQPPGS
AGIEHLRVLDLEPGKRVQAVDQVTVVHLVAYAGGSTWRSAAATVPAERVNAVGMIRD
LVAALRABGPAPVLLPASTTQANPAAPSAOAKIEARILRQATGVDGVILR
LPAIYHSGSGSQGRGVTVAMIRALAGEPIITWHEGSVRRNLLHVEDVATATAAL
HNHEALVGDVWTPSADARPGRIETVAASVARQTNPAVPVSVVPPENAEANDFR
SDDFDSFEFRLTGMHPRVPLAEGIDRTVAALISTKE"
12181..22821
/gene="megAI"
CDS
12181..22821
/gene="megAI"
/translation="polyketide synthase"
Query Match 4.2%; Score 21; DB 1; Length 47981;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 124 TTCAGCTCGTGGCGCGGTC 144
Db 22447 TTCAGCTCGTGGCGCGGTC 22467

RESULT 27
AX112026 AX112026 47981 bp DNA linear PAT 01-MAY-2001
LOCUS Sequence 1 from Patent WO0127284.
DEFINITION
ACCESSION AX112026
VERSION AX112026.1 GI:13938925
KEYWORDS
SOURCE Micromonospora megalomicea
ORGANISM Micromonospora megalomicea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE
1 McDaniel, R. and Volchegurkey, Y.
AUTHORS Recombinant megalomicin biosynthetic genes and uses thereof
TITLE Patent: WO 0127284-A 19-APR-2001;
JOURNAL Kosan Biosciences, Inc. (US)
FEATURES
source
1. 47981
/organism="Micromonospora megalomicea"
/mol_type="unassigned DNA"
/db_xref="taxon:47868"
complement(1..144)
/note="unnamed protein product; megBVI (megT),
TDP-4-keto-6-deoxyglucose-2,3-dehydratase; SEQ ID NO: 2=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37805.1"
/db_xref="GI:13938926"
/db_xref="REMTREMBL:CAC37805"
/translation="MGDRVNGHATPESTQSAIRFLTRHGGPPPTATDDVHDWLAHRAAE
HRLR"
928..2061
/note="unnamed protein product; megDVI,
CDS
```

```
TDP-4-keto-6-deoxyglucose 3,4-isomerase,
TDP-4-keto-6-deoxyhexose 3,4-isomerase; SEQ ID NO: 3=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37806.1"
/db_xref="GI:13938927"
/db_xref="REMTREMBL:CAC37806"
/translation="MAVGDRLRLGRELQMARGLYMGFGANGDLYLSMLSGRDDDPWTW
YERLRAAGRPYASRAGTGVGDHRTAEVLIADPGFTGPPDAARMWMOVAHCPAASWA
GPFREFYARTEDAASTVVDADWLQRCARLVTELGSRPDLVNDFAREVPLALCTAPA
LKGVDPDLRSEWTSATRVCLDQAQVSPQOLATEQALTDDEIDATGGRDAALVGVV
AELAANTGVNAVLAVTELPELAARLADDDPETATRVVTSRTSPGVHLERTTAASDRR
VGGVDVPTGGVTVVAAANRDPEDVTDPRDFVDVDRGGDAEILSRSPGSPRTDLALV
ATLATAALRAAAPVLRPLRSRGPVIRRRSRPVARGLSRCPVEL"
2072..3382
/note="unnamed protein product; megDI, rhodosaminyl
transferase (eryCIII homolog), TDP-megosamine
glycosyltransferase; SEQ ID NO: 4= translated amino acid
sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37807.1"
/db_xref="GI:13938928"
/db_xref="REMTREMBL:CAC37807"
/translation="MRVVFSSMAVNSHLFGLVPLASAFQAAAGHEVRVVASPALTDVDT
GAGTAVPVGDDVELVEWHAGQDIYVMTLTDWQSDHTMSDDLLGMOTTTFTPT
PFALMSPDSLIDGMVERCWRPDIWVEPLTFAPIAARVTGTGPHARMLMGPDVATR
ARQSFLLAHQVEHREDPLAENFDWTLRRFGDDPHLSFDEELVLGQWTVDPPIEPL
RIDTGVTVGVRYVYNGSPVPAWLLKREPPRRVCLTLGSSREHGGISGEMLD
AIADIDAEFVATFDQQLVGVSPANVTAGVPMNVLLPTCAATVHGGTSGMLTA
AIHGVPQIILSDADTEVHAQQLDGLAGLSLPAVAGMTAEHLRGAIERVLDEPAFVRLGA
ERMDDGRTDPSPAQVVICQDLAADRAARCRQPRRTAEPLPR"
3462..4634
/note="unnamed protein product; megG (megY), mycarosyl
acyltransferase, mycarose O-acyltransferase; SEQ ID NO: 5=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37808.1"
/db_xref="GI:13938929"
/db_xref="REMTREMBL:CAC37808"
/translation="MVTSTNLDTTARPALNSLTGMRFVAAFLVFFTHVLSRLIPNSYV
YADGLDAFWOTTRGVSGFFFLSGVLTWGSARSDSVSWFRRRVCKLFPNHLVTAF
AAVLFLDTGOAVSSEALIPNLLIHAWFPALBISFGINPVSWSLACAEFFVLCPLF
LFWISGIPERLWAAVFAIWAQVPAVADLLPSSPLIPGLSEYSAIQWFLYTFP
ATRSDEFILGILAILITGRWINVGLLPAVLLPFFVFAVSLFLGVTAISSMWLLP
LVLIASGATADLQOKRTFMNRVMVGLDVSFALYVMVFLVIVYGADLLGFSQTEDA
PLGLALPMIIPFLAVSLVSLWLLXRFVLPVNRNWARPASARRKATEPEQTPSR"
4651..5775
/note="unnamed protein product; megDII, deoxysugar
transaminase (eryCI, DnrJ homolog),
TDP-3-keto-6-deoxyhexose 3-aminotransaminase; SEQ ID NO:
6= translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37809.1"
/db_xref="GI:13938930"
/db_xref="REMTREMBL:CAC37809"
/translation="MTTVVSYLLEYERERADILDVOKVPASGSLIIGQSVENPETE
YARYHGIAHCVGVDNGTNAVLALESVGGVRDDEVVTNTAAPTVLAIDEGARPVF
YVRDEYLDMDTLVEAAIPRTKAIVPVHLVGCDDMTALRELARGLKLVEDCAQ
AHGARRCRLAGTMSDRAAFSFPYTKVLGAYDGGAVTNDDETARALRLRYGMEE
VYVYTRFCHNSRLDEVQAEILRKRLTDLDAYVAGRAVQAQRYVHTMTGFAHLGVASG
PVWTDGNEHFVYVYVHRPRDEILKLRDGDVLSINISYVWPVHTMTGFAHLGVASG
SLPVTRELAGEIFSLPMVPSPLHQDRVIEAVREVITGL"
5822..6595
/note="unnamed protein product; megDIII,
daunosaminyl-N,N-dimethyltransferase (eryCVI homolog); SEQ
ID NO: 7= translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37810.1"
CDS
```



```
/db_xref="GI:13938931"
/transl_table=11
/translation="MRVSHSTTSSTSDVAVPVERADIYVHDFVHGRKGKGYRAEADALVBA
RHTFPAAILLVACCTGSHVLELADSPREVGVGDLISAAMLATPAANDPGRELHQGDM
RPFSLDRREDVTCFSSGTYLVDIELDRANVANLAGHAPGGTLVVEPWPFETPRP
GWGADLVTSGGRRISRMGSHTPAGUPDRTASRMTIHYTVGSPGAEIHFTTEVHVMTL
FARAAYEQAFQAGLSCSVGHDLFGSLFVGVAAPGR"
6592..7197
/notes="unnamed protein product; megDIV,
TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVI1, dmnu
homolog), TDP-4-keto-6-deoxyhexose 3,5-epimerase; SEQ ID
NO: 8= translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37811.1"
/db_xref="GI:13938932"
/db_xref="REMTREMBL:CAC37811"
/translation="MRVEIGIEGVFTFTPTFADRGVGTAYQEDVFVAALGRPLF
PVAQVSTRSGVGVGVRGHTTTPGSMKATYVCARGAMDFAVDIIRPGSTFGRAEPV
ELSAEWMVGLYLPVGMGHLFVSLDDTTLLVYLSAGVYPDKERAVHPLDPELALPIPA
DLPLVMSERDRVAPTLEARDQGLIPDYAACRAAAHVVRT"
7220..8206
/notes="unnamed protein product; megDV, TDP-hexose
4-ketoreductase (eryBIV, dmnuv homolog),
TDP-4-keto-6-deoxyhexose 4-ketoreductase; SEQ ID NO: 9=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37812.1"
/db_xref="GI:13938933"
/db_xref="REMTREMBL:CAC37812"
/translation="MVVLGASGELGSAVTHALADLPVRVRLVARREVVVPSGAVDYE
THVDLTPGALAEVADARAVFPFAAQIRGTGWRISDDVVAERTNGLVRLDIAY
LSRSHAPVVPSCNVQGRVTAQVIGDSQODHPGVYDROKHTGEQLLKQATAY
AIRATSLRLPVGVRAGTADRGVWSTMIRALCQPLTMWHDGVRELLYVTD
AARFVTDALPDAGALGELGLTGRSWPLGVFOAVRSVARHTGEDVPVVSVPFA
HMPDPSLRSEVDPARTVATVGWRTVTAEVADRTVAALAPRRAAAPSFS"
complement(8228..9220)
/notes="unnamed protein product; megBII-1(megDVII),
TDP-4-keto-L-6-deoxy-hexose 2,3-reductase; SEQ ID NO: 10=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37813.1"
/db_xref="GI:13938934"
/db_xref="REMTREMBL:CAC37813"
/translation="MGTGTGASRVVRGSRNLHTSRNLGTVNFSGRVTDLRLMD
HALRGVNCIDTIYGNRLYKHTHELGVGFAQGGRRRETVIATKVGSESRVN
DGLSARHVAACNSLRRLGVHDIIYQTHIDRAAPWDEWQAHLVSGKGVY
GSSNLGWHIAAQAESARRNLLGMIHQCLYNLAHRPELDVLPAAQAGVGVFAWS
PLHGGLSGVLEKLAAGTAVKSAQRAQLLPAVRPLVEAYEDYCRILGADPABVGLA
WVLSRGLGAVIGPTPEQLDSALRAELTLGEELELEAIFPAPAVDGEV"
complement(9226..10479)
/notes="unnamed protein product; megBV, mycarosyl
transferase, mycarose glycosyltransferase; SEQ ID NO: 11=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37814.1"
/db_xref="GI:13938935"
/db_xref="REMTREMBL:CAC37814"
/translation="MRVLLTSFAHRTFQGLVPLAWALHTAGHDVVRVASQPELTDVVV
GAGTSLVPLGSDHLRFDISPEAAQVHRVYTTDLDFARRGPRLSWEFLHGEATSRF
VFPVNNDSFYDELVEFAMDREPLVLWEPTTFAGAVAKACAGAHAPLLWGSDLTGY
FRSRQDLRGQRADRDPLGLGMLTEVAGRFGLDYSDDLAVGQSWDQLESFRLT
GLESVHTLTPYSGSVVFWMLRSDGVRCVFTGGYSALGITSNPQFELTLATLAR
FDGEIVTVRSGDLPASPDNVLVDVFPNMNLLPFCRAVHHGGAGSWATLHGVQP
ISVAHEMPCVLRQTEBLGAGVFLRPDEVDATIMQALATVVEDRSHAENAEKLQRE
ALAAPTPAEVVPVLEALAHQHRDR"
complement(10483..11424)
/notes="unnamed protein product; megBIV, TDP-hexose
4-ketoreductase, TDP-4-keto-6-deoxyhexose 4-ketoreductase;
SEQ ID NO: 12= translated amino acid sequence"
```

```
/codon_start=1
/transl_table=11
/protein_id="CAC37815.1"
/db_xref="GI:13938936"
/db_xref="REMTREMBL:CAC37815"
/translation="MTRHVTLLGVSGFVGSALLREFTTHPLRLRVARVARTGSRDPPGS
AGIEHURVLLLEFPGRVAVQVADTDVVVHLVAVAGSTWRSATVPEASRVNAGINRD
LVAALRPGPAPVLLFASTTAANPAAPSAQHKIEARILDRQATEDGVVDGVLIR
LPAIYGHSPGSGQTGRVVTAMIRALAGEPITMMHEGSRVRLNLEVEDVATATAAL
HNHEALVGDVWTPSDEARPLGEIFETVAASVARQTGNPAVPVVPVPPENAEANDFR
SDDFOSTERTLTGWHPRVPLAEGIDRTVAALISTKE"
12181..122821
/notes="unnamed protein product; megAI; SEQ ID NO: 13=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37816.1"
/db_xref="GI:13938937"
/db_xref="REMTREMBL:CAC37816"
/translation="WYVDPLLGTRTHPGPLFPFNPWPLCGHNEPELRARARQLHAYLE
GISEDVAVGAALARETRAQDQPHRAVVVASSVTETLTAALAAQAQRPHSPVVRGVA
RPTAPVVFVLPQGQAWPGMATLLAESPVFAAAMRACERAFDEVTDSLTELDSPF"

Query Match 4.2%; Score 21; DB 6; Length 47981;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 TTGACTCGCTGGCGGGTC 144
|||||
DB 22447 TTGACTCGCTGGCGGGTC 22467

RESULT 28
AY310323/c
LOCUS 138203 bp DNA linear BCT 06-DEC-2003
DEFINITION Streptomyces sp. FR-008 heptaene macrolide complex synthesis gene
cluster, complete sequence.
ACCESSION AY310323
VERSION AY310323.1 GI:34766435
KEYWORDS Streptomyces sp. FR-008
SOURCE Streptomyces sp. FR-008
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1 (bases 1 to 138203)
AUTHORS Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K.J., Lee, S.Y. and
Deng, Z.
TITLE Organizational and Mutational Analysis of a Complete
FR-008/Candididin Gene Cluster Encoding a Structurally Related
Polyene Complex
JOURNAL Chem. Biol. 10 (11), 1065-1076 (2003)
PUBMED 14652074
REFERENCE 2 (bases 1 to 138203)
AUTHORS Chen, S., Huang, X., Zhou, X., He, J., Bai, L., Jeong, K.J., Lee, S.Y. and
Deng, Z.
TITLE Direct Submission
JOURNAL Submitted (30-MAY-2003) Bio-X Life Science Research Center,
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,
China
FEATURES
source
1..138203
/organism="Streptomyces sp. FR-008"
/mol_type="genomic DNA"
/db_xref="taxon:206662"
/complement(574..1950)
/gene="fscO"
/complement(574..1950)
/gene="fscO"
/notes="putative FAD-dependent monooxygenase"
/codon_start=1
/transl_table=11
/product="FscO"
/protein_id="AAQ82549.1"
```

```
/db_xref="GI:34766436"
/translation="MPSTCAPTSGIGRGCPHRRKGTSGRRITAVTPRCGKVKGH
QHAQKRVSRMGSGVAVVGGSIAGCAALASRGAEKITVFERADARLDRGVGIGV
HDARYAELRDAGYNADMEWAPLNRVWSVRDGAEGHRTGTQFPFRAYNWGLWS
ELRRVPTATYRAGAEVAEQDAGVTVRLADGEPRFDLVIQDGYRVSREAY
YCAAGVAGYTGMRGTPSDVSGPLSDGIDAHNITFPGHCMAYRIPDGGSGHRLNWVL
PTAPRIGLHPDLRTPTSLPPGRINAELTEHLRALVAEHPPTFWAAKLLSTPAETTP
IQPYLDYDHYATDRAVULIGDAASVAPPHIGAGSVKALODATALEAAWAGESKKEI
LEGTHAARGVGTAMVGLARMGSAVEETPDWKSAMQAEFDAMWQEQNGSDRRSGF
GGHSIKSR"
complement (2264..3037)
/gene="pabc"
complement (2264..3037)
/gene="pabc"
/notes="putative 4-amino-4-deoxychorismate lyase"
/codon_start=1
/transl_table=11
/product="Pabc"
/protein_id="AAQ82550.1"
/db_xref="GI:34766437"
/translation="MIEDGEPAGPEALASLALTYGHFTTLLVNGEVRGGLDLHLER
LIRDCRTLDALDPDRVKLRARAAPTDGRATVAVTVDFPALNIGNTAADAPGILV
TSRPAPDPGPKVRVSRHREDLPEVKSGLCPTLRRLRQAQAGYVDFLTPGDGD
ILEGTMVGVIRGDEVVPGVEVLGTTQRLRRGRRSHRTGPRGPRFLSRGGLRD
QRRVGVPTGIDDRPEFAAHSVTRLAEIYQALPGSPL"
complement (3150..3818)
/gene="fscRI"
complement (3150..3818)
/gene="fscRI"
/notes="putative transcriptional activator"
/codon_start=1
/transl_table=11
/product="FscRI"
/protein_id="AAQ82551.1"
/db_xref="GI:34766438"
/translation="MDPAPADPAGSPERRAAVLADGFRAGAYVACLDPESLTIQVNWQ
EFERFGSSSELCSGRFCDLIVHPSVQOPLMHQFARMLDGKRHRATEVIAVDORTA
STLPNALAVGGPTPDVAAITLVNMAAEEEGADWAPRKLLSEIDARLTLEGIAA
GVSTPLASRLYLSRQGVYHVTGLRKLKVENRAALVSRAYSVMGVLVKGTWPPKVVZ
DFIK"
complement (4377..7205)
/gene="fscRII"
complement (4377..7205)
/gene="fscRII"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="FscRII"
/protein_id="AAQ82552.1"
/db_xref="GI:34766439"
/translation="MLERTTIALVDALGAATGRSSLVLITQPLGIGRSALLQRL
SLADDDRVIRLHAAPMEQDFAGVVRQLFETLLGDSPEDARDMDHASFARH
VLADDAAPGDAQIAATEAVLHGLSLANVASDRLLILVDDIQWSDVPSLRWLT
LAKRLHLAVVVCALRDGPRSHHTLAREIREAGTQTLRPASLSLTAHELVRHFH
EAGDDEFVQACHEASVGNPLFLSLIVGTGFLRRPLAEHASTARRLRPSQLRKLAS
ILRTQAPVYDLAAIILIGEQSDAPTARLAGLSIGVAGALRALGALGAADPEP
RTIHSVRDAESTLTMVOREKHDEAALLVAAGSPAQVAAQLMAVTVRQPVAVD
VLRAAADTRALRGADTAAGYLCRALDSPAAGVGRLLVELGTAERGFDPACERH
LQAQMTLLPDRADVAALURI SPTALGPAPLAVDLLRQAEDLGPADGLTGTDRDLA
LLEALRLHCEHEDPHELAASARHLKELGPEPVDGTGSGRELTIALGAAITGLCHPLA
SVVRLAGVLDRPATSARHISTPLAIIITMAADSVEAVDSWLAVERANERGTTTP
ANALVHEQALVHLCGRLAQBOSETALGMEPAGHQLDAAWTHAIVALESRDP
SLAEVIRKASRIROPGLTILGLILLEATREIAAGDQGTALDSLHSCGRQLEAGWH
NPVILFPWRPWAHVHRRLDGIDTARSLABEEYARAGQWAPGVGAAATLLRGLDEGG
RGTLGLREAVTVLRGSANQLEALSTLAERLGGGAETALAREATLATSCGAPGL
ACRAAGAGHRSGPPATPEDLVTPERRVTVGLACRLITQETAGALGVSSRAVEKHLT
HAYRILIGSRRELIVLFSRAEYAGTD"
complement (7210..10254)
/gene="fscRIII"
complement (7210..10254)
/gene="fscRIII"
/notes="putative transcriptional regulator"
/codon_start=1
```

gene

CDS

gene

CDS

gene

CDS

gene

CDS

```
/transl_table=11
/product="FscRII"
/protein_id="AAQ82553.1"
/db_xref="GI:34766440"
/translation="MYDLHMPPIELPLLRERDELAALSAVIGELSGRPAVVVTGEP
EQGGLDLRWAAAYATDAGRLVLSAHATPAEHYRYGVVQALLAEENRALPRFLTLD
BQGLGLINGLAINAVDRPTLLVVRDQMLDPSLRMLELALRLRPAPVALLTSTT
GTAITRPEWSVGTPLTGLATTIELALPGLTSGGTATVLRAPGAPDPAPTEALAEAT
RGIAVVDHDLDRFARAGYSPADRIODPLRALTAEVVGCDHATRALSDLRDPAADVLR
ALAVCGDLIDPLVCTIAGHPSVSESLRAALAASGLTLLRDRHPRVQDAAVRAVLE
EMPAADRAELYARAAGLAQRAADDQGIADLILLARVPDGFWDVILRRGTGTSALRG
RRDLAVAYLARALDEPLAEADRIEFQLASVEMVMTAFAERERLGLLIRATRPQPGA
GLRARATDCLLGGTTRAARHALAGIDSAPAEPRRGPTGAPAVPEAPEASTPTT
HLVPHRPVAGTPPGATGDRRSAPDLPTGTPRAHDAPSGPGEAAGSDPTAGDPEDP
AVERHELTVLCRVASFRLRHDDADLDVLPALPDVAVHNPVAVWMEHGVAGVHRE
EVRRLASAFVPAANTGRPLLVVRLGCRALLTDDGDEAENHGLGALLDESHERRAAL
SVAHILTVRAELHHRGRPDAAADLAAQAELEPLDLRLHPLFLPTWLALSMITDLQNG
HTERARETAERPLPLAHESATTAQLFARGVLTARTDDDEHOAREHFRACRWLLRHG
CANPAVQWPSRLAAEAHTLGDTEEARLVHEVRLARRWGAAPLGAQSLAVVTE
ENRVENLRAAVTTLSAASARTAYTRAVTELAAREENERRTAVAGLGAATVLPVAP
PPGPHVPSAVARPPGPRPAVLTGRRTAKAPGDAAVHGLTAVERTAAALAAQGLNRR
BIATELATTVRAVELRSLGVRKLRIRGORELRALVQEAEGS"
complement (10298..13315)
/gene="fscRIV"
complement (10298..13315)
/gene="fscRIV"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="FscRIV"
/protein_id="AAQ82554.1"
/db_xref="GI:34766441"
/translation="MWLRWAGDRSGAAGPSGASPPSLAPSDARESEVPHRSRAVP
PAPPLVGRHAQDLVLARHADARAGRPRLVLDGPKGTALRLAALAEDEGLAGMT
TLYGSCRAVDATGYSVGRALFGGLVLTGKRTGSPFLVVGARRALPALAADPGELDA
DPGFTSVQLGYLANLMAADGVLVLDVHWCDESLRMLDLELLRAGDGLPLLV
LAVHTGSPLEIVLRLDLADLIVAHPLPASLGLPLDTAEVLAELTAHSPQDPLSPFVRQ
LSRTGSGPLEIVLRLDLADLIVAHPLPASLGLPLDTAEVLAELTAHSPQDPLSPFVRQ
VARAVLVDGSDRTVLAALGRVSVVVEAEVILRDAGLIHPGHLESHDLVRAAVLID
AVGASGVAVLRRAARLLSDVGRSPEETATQLLLVPGTPDDVAVSRDAAQAQORG
ACEAAALERVRAKDPDVLRLGKALAEPTDPAFVTLLEHSHLTDIVRAAAT
AVQLGLTCLAVQSPDGARLAEALDADLAELGPEPEPADRELRTTESALLIVGSDDE
KATLPIILRTQEGTTPQDTPAQOQLAMLSVLSAEGGGAETVGRARRALRPGV
PLGVMSLAPTSIALADENEAEBEVLTGSGDTPAAVTYVLAELSTRSLFRLENG
AVPDMAADQTALEILGQERWGDATMPHTAASVLTREGDPGRALAEALDKRPHLD
RFVWEIHTLMARGALAADGLDQALQVFGSGGASMAEAGLTNPNVLAPEWLEACLL
CEAGREBAGRAAAGHTLAERWGTTRALGAAALARGAARAGTARTGALUREAVALLAD
SPARSGHARGILLGRALVEDGAVREGRHREAVLGRRCGCVALARRADELTAAG
GRMREVTASPLDMLTGTTERTVAGLVASGAGNEVAESLFVTVTVTVELHLTSVYKLGIV
ARRGDLTALREAGATARPAAERRPPHAKRRNP"
13522..14898
/gene="fscMI"
13522..14898
/gene="fscMI"
/notes="glycosyltransferase responsible for mycosamine
attachment"
/codon_start=1
/transl_table=11
/product="FscMI"
/protein_id="AAQ82555.1"
/db_xref="GI:34766442"
/translation="MDSARPILFVSLPESGLNPLLVLGALARRRGVLDLWFATDEH
RREEVEALSDVSKVSFASIGEVVPELSAVTWDEVVREVTQSSRFKAHVRQSVRP
ALQCARILEQVVDVDEVRPALMVVDCVAGTVDLALARGIPVYLVNVPFVASNVLTSHN
PGASVYTPKFPVPSGLPSPARMVSKQKLNTKWLKTFGLHFDHPMAALSRDAALRK
ELIGAPYPPNAPRVDEAAAVVCSVAELDYPFDIPDRVSLVAGVLPPLPQAPDDDEVTR
WLDQSSVVMYMGFTITRLTREVAALVEVARMSTGHQFLWKLKPEQOHLPEAGSL
PDLRVESVPSQDLVLAHPNVSVFSGHGGNAVHEGVYFGKPVVVRPVPVDCFDQAV
RGDRFGISLTLDKPHVDPDDVDKLTITVTSIDPAFTEARLGLAQRAAGRAAADL
VTGLLPPAA"
14953..16011
/gene="fscMII"
14953..16011
```

gene

CDS

gene

CDS

```

/genes="fscM11"
/notes="putative aminotransferase responsible for
mycosamine biosynthesis"
/codon_start=1
/transl_table=11
/product="FscM11"
/protein_id="AA082556.1"
/db_xref="GI:34766443"
/translation="MAFTHPVSPALDGRLEVVDVSGWISSGCPYVRRFEAPFA
RWNGVAGVACSGTALIALRALNIGPDEVIVFETNVASAMAVTYTGAPFVVD
CGDMDINDVIRIEKIPTRVAVHVGVRCDMDALQCNLRVVDVDSAEHAG
VREVGDIACSLFANKIITRAGEGVCILTDPRLEAQHLRAWAFTRDSFLHKAHG
NRYMTOAQVALDEILATREIEARYDAGLQDLPITLMPARDVLWYDLR
AEERELRAHLDAEGTRELTRFFKPMSPQGYLDFVWPTLNARFSDGLYLPHTGILT
AAQDEYITGAVRDFRAG"

```

Query Match 4.2%; Score 21; DB 1; Length 138203;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GGCTTCGACTCGTGGCGCG 141

Db 40157 GGCTTCGACTCGTGGCGCG 40137

RESULT 29  
 AK110400/c  
 LOCUS  
 DEFINITION Oriza sativa (japonica cultivar-group) cDNA clone:002-165-F08, full insert sequence.

ACCESSION AK110400  
 VERSION AK110400.1 GI:32995609  
 KEYWORDS FLI CDNA; oligo capping  
 SOURCE Oriza sativa (japonica cultivar-group)  
 ORGANISM Oriza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1  
 AUTHORS The Rice Full-Length cDNA Consortium, National Institute of  
 Agrobiological Sciences Rice Full-Length cDNA Project Team;  
 Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,  
 Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,  
 Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,  
 Ohtsuki, K., Shishiki, T., Foundation of Advancement of International  
 Science Genome Sequencing & Analysis Group; Otono, Y., Murakami, K.,  
 Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,  
 Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,  
 Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,  
 Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimizu, A., Miura, J.,  
 Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN;  
 Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,  
 Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,  
 Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,  
 Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Yoshino, M., and Hayashizaki, Y.  
 Collection, mapping, and annotation of over 28,000 cDNA clones from  
 japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)  
 MEDLINE 22752273  
 PUBMED 12869764

REFERENCE 2 (bases 1 to 3161)  
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,  
 Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,  
 Hori, F., Hotta, I., Iida, Y., Ikeda, R., Imamura, K.,  
 Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,  
 Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,  
 Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,  
 Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
 Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,  
 Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,  
 Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,

Namiki, T., Narikawa, R., Niikura, J., Nishii, K., Nomura, K.,  
 Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,  
 Osato, N., Ota, Y., Otono, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,  
 Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,  
 Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,  
 Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,  
 Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,  
 Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,  
 Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and  
 Yoshimura, A.

Direct Submission  
 Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of  
 Agrobiological Sciences, Department of Molecular Genetics, Head of  
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki  
 305-8602, Japan (E-mail: shikuchi@nias.affrc.go.jp,  
 Tel: 81-29-838-7007, Fax: 81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica

rice.

URL : http://cdna01.dna.affrc.go.jp/cDNA/  
 NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,  
 Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,  
 Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,  
 Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and  
 Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Otono, Y., Iida, Y.,  
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,  
 Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,  
 Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,  
 Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,  
 Yoshimura, A., Matsubara, K. and Murakami, K.  
 Genome Exploration Research Group in Riken Genomic Sciences Center  
 and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,  
 Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,  
 Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,  
 Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,  
 Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,  
 Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,  
 Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,  
 Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,  
 Yasunishi, A. and Hayashizaki, Y.

FEATURES  
 source

1..3161  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="002-165-F08"

ORIGIN

Query Match 4.0%; Score 20; DB 8; Length 3161;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 TGCTCGGCTCGCCGACACC 378

Db 2126 TGCTCGGCTCGCCGACACC 2107

RESULT 30

AK112081

LOCUS

DEFINITION

Oriza sativa (japonica cultivar-group) cDNA clone:001-124-H09, full insert sequence.

AK112081

ACCESSION

VERSION

KEYWORDS

ORIGIN

ORGANISM

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;

3835 bp mRNA linear PLN 29-OCT-2003  
 Oriza sativa (japonica cultivar-group) cDNA clone:001-124-H09, full insert sequence.

AK112081.1 GI:37988744

FLI CDNA; oligo capping.

Oriza sativa (japonica cultivar-group)

Oriza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

# REFERENCE AUTHORS

The Rice Full-length cDNA Consortium, National Institute of  
Agrobiological Sciences Rice Full-length cDNA Project Team:  
Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,  
Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,  
Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,  
Ohtsuki,K., Shiehiki,T., Foundation of Advancement of International  
Science Genome Sequencing & Analysis Group, Ootomo,Y., Murakami,K.,  
Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,  
Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,  
Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,  
Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,  
Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:  
Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,  
Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,  
Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,  
Saito,R., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T.,  
Yoshino,M. and Hayashizaki,Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from  
japonica rice

Science 301 (5631), 376-379 (2003)

22752273

12869764

# REFERENCE AUTHORS

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,  
Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,  
Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,  
Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,  
Kanagawa,S., Katoch,H., Kawagashira,N., Kawai,J., Kawamata,M.,  
Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,  
Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,  
Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,  
Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakahama,Y.,  
Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K.,  
Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M.,  
Ooka,H., Osato,N., Ota,Y., Otomo,Y., Rydu,R., Satoh,H., Sakai,C.,  
Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K.,  
Shibata,K., Shingawa,A., Shiraki,T., Shishiki,T., Sogabe,Y.,  
Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M.,  
Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K.,  
Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J.,  
Yokomizo,S. and Yoshimura,A.

Rice full-length cDNA

Unpublished

3 (bases 1 to 3835)

Kikuchi,S.

# REFERENCE AUTHORS

Submitted (12-SEP-2002) Shoshi Kikuchi, National Institute of  
Agrobiological Sciences, Department of Molecular Genetics, Head of  
Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki  
305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,  
Tel:81-29-838-7007, Fax:81-29-838-7007)

This clone is one of the 32K full-length cDNA clones from japonica  
rice.

URL : <http://cdna01.dna.affrc.go.jp/cDNA/>

NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K.,  
Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,  
Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,  
Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shiehiki,T.,  
Yamamoto,M. and Nakahama,Y.

FAIS Genome Sequencing & Analysis Group: Ootomo,Y., Iida,Y.,

Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M.,

Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M.,

Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M.,

Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M.,

Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.

Genome Exploration Research Group in Riken Genomic Sciences Center  
and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,

Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,  
Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,  
Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,  
Itoh,M., Kagawa,I., Kanagawa,S., Katoch,H., Kawai,J., Kouda,M.,  
Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N.,  
Ota,Y., Satoh,H., Sakai,C., Sakai,K.,  
Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K.,  
Shingawa,A., Shiraki,T., Sogabe,Y., Tagami,M.,  
Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,  
Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,  
Yasunishi,A. and Hayashizaki,Y.

# FEATURES

## source

1..3835  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="001-124-H09"

## ORIGIN

Query Match 4.0%; Score 20; DB 8; Length 3835;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 TGCTCGGCTCGCGACACC 378

|||||

Db 722 TGCTCGGCTCGCGACACC 741

|||||

RESULT 31

AF262754

LOCUS

AF262754 3900 bp DNA linear BCT 27-SEP-2000  
Amycolatopsis mediterranei DSM 46095 rifamycin biosynthetic gene  
cluster, partial sequence.

ACCESSION

AF262754

AF262754.1 GI:10312147

KEYWORDS

ORGANISM

Amycolatopsis mediterranei

Amycolatopsis mediterranei

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Pseudonocardiaceae; Pseudonocardiaceae; Amycolatopsis.

REFERENCE

1 (bases 1 to 3900)

Lal,R. and Kaur,H.

Cloning and partial characterization of presumptive rifamycin

biosynthetic gene cluster from the actinomycete Amycolatopsis

mediterranei DSM 46095

Microbiol. Res. (2000) In press

REFERENCE

2 (bases 1 to 3900)

Lal,R. and Kaur,H.

Direct Submission

TITLE

JOURNAL

Submitted (02-MAY-2000) Zoology, University of Delhi, Delhi 110007,

India

FEATURES

source

1..3900

/organism="Amycolatopsis mediterranei"

/mol\_type="genomic DNA"

/strain="DSM 46095"

/db\_xref="taxon:33910"

106..>3900

/note="similar to Amycolatopsis mediterranei S699

rifamycin biosynthetic gene cluster; contains module 2

ACP, all of module 3 including KS, AT, and ACP, as well as

part of module 4"

/codon\_start=1

/transl\_table=11

/product="rifamycin polyketide synthase"

/protein\_id="RAG16689.1"

/db\_xref="GI:10312148"

LDDAGGPDAAFFGSIQSPNEALMDPQRLLETSGLFERAIDPTTLAQDQIGVFG

```

NSHDYWRTHGAAGVGRFLRTGASAVLSGRVAYHFGVEGPAVTVDTACSSSLVALHM
AVOALQCEGAMALAGVMVGTVEFVFSRQKGLAPGRCFKAFADGADGTGWSGV
GLLVLRLSARRRGGHOVLAVRGSVNSDGSNGLTAPNGPAQQRVIRKALAAAGLS
PSDVAVEAHGTGTTGDPLEAEALATATYQDQDEQPLNLSGVKSNLGHQTAAGVAGV
IKVMWRRGVLPTRTHVDPRSSHVDWSAGAVELLTEARDWTSQDPRRAGVSSFGIG
GTNAVVLLEVPAPVATPRPEPAGFLVPVLLSARSAGLRQGAARLAFGAEVRVPD
VAHALATTRAQLDRAVTLASDRDLRADLAAPFGVVTGTPTVDGLAVLFTQCGSQM
AGMRELAGTFPVRNFAEAACHAVDHLGRAPVREVVDFDRLDQTMVTLQCGALFPAV
ETALFLFASGVRPDLLAGHSIGELAAAHVSGVLDLAEAGELVAARGRLMOALPAGG
MVAVOAAEDVAPLDGTVCVAANVGPDVSLSGAEDAVLAAGLAARGRTKRLP
VSAFSPMLMEQMLGDPRVAVERLTTRAGTLVPVSTLTGKLAALDPGYVVRQVRDAV
RFGDVAELNAGATTYLEGGGALGAALMTLGGPEQSCVATLNRKNAGAEVPGVITA
LAEHLVGRVTDVMTALLDRTGATVTLPTTAFQHERYWGTDTPSPASPVAPARPAE
PVDHVPALVRESVAVLGRDTPAADLDRFSKDLGSLGAIKLNRRLDFTGVDPLPS
TLIFDVPNAAVLADHLRQWDLGERPAVAAVPRVDSEPTAI VGMSTRLPFGADSPBE
LWNLVARGDRAISGFPVDRGDLGLVHPDPAHGTSTYRTGGFLHDAADAGLFGI
SPREALMDQOQRLLETSWEALERAGVDPLSARGSDVGVFTGTIVHHDYVTRREVP
DVQGYTMTGTASSVASGRVAYVFGFEGPAVTVDTACSSSLVAIHLAAQALRQGECSNA
LAGAGTWMASPDAPLEFSRQGLSADGRCKRAYAGADGTGWAEGVGVVLERLSVARE
RGH"

ORIGIN
Query Match 4.0%; Score 20; DB 1; Length 3900;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ACGTGATCTCGACTACCC 212
|||||
3022 ACGTGATCTCGACTACCC 3041

RESULT 32
AKI21779
LOCUS Oryza sativa (japonica cultivar-group) cDNA clone: J033093K02, full
insert sequence.
DEFINITION
AKI21779
ACCESSION
AKI21779.1 GI:37991402
VERSION
AKI21779.1
KEYWORDS
F11 cDNA; CAP trapper.
SOURCE
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1
AUTHORS
The Rice Full-Length cDNA Consortium, National Institute of
Agrbiological Sciences Rice Full-Length cDNA Project Team;
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
Ohtsuki, K., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Ootomo, Y., Tsunoda, Y.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M., and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
JOURNAL Science 301 (5631), 376-379 (2003)
MEDLINE 22752273
PUBMED 12869764
REFERENCE
2
AUTHORS
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
Hori, F., Hotta, I., Iida, Y., Ikeda, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
Kana
```

```

Kana gawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodana, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, Y.,
Nakamura, M., Namiki, T., Narikawa, R., Niikura, J., Nishii, K.,
Nomura, K., Numasaki, R., Ohneda, E., Ooka, H., Ohtsuki, K., Oka, M.,
Ooka, H., Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C.,
Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K.,
Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y.,
Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M.,
Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K.,
Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J.,
Yokomizo, S. and Yoshimura, A.
Collection, mapping, and annotation of 28K full-length cDNA clones
from japonica rice
Unpublished
3 (bases 1 to 4116)
Kikuchi, S.
Direct Submission
Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of
Agrbiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp)
Tel: 81-29-838-7007, Fax: 81-29-838-7007
This clone is one of the 32K full-length cDNA clones from japonica
rice.
URL : http://cdna01.dna.affrc.go.jp/cdna/
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,
Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T.,
Yamamoto, M. and Nakahama, Y.
FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M.,
Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M.,
Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Oka, M.,
Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M.,
Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishii, K., Nomura, K., Numasaki, R., Osato, N.,
Ota, Y., Saitoh, H., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M.,
Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A. and Hayashizaki, Y.
Location/Qualifiers
1..4116
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/db_xref="taxon:39947"
/clone="J033093K02"
FEATURES
source
1..4116
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/db_xref="taxon:39947"
/clone="J033093K02"
ORIGIN
Query Match 4.0%; Score 20; DB 8; Length 4116;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 TGCTCGGCGCTCGCGACACC 378
|||||
Db 1011 TGCTCGGCGCTCGCGACACC 1030
```

```
RESULT 33
AF305325 7080 bp DNA linear BCT 18-JUL-2001
LOCUS Comamonas testosteroni strain BR6020 protocatechuate meta
DEFINITION dioxygenase locus, complete sequence.
ACCESSION AF305325
VERSION AF305325.1 GI:14861212
KEYWORDS Comamonas testosteroni
SOURCE Comamonas testosteroni
ORGANISM Comamonas testosteroni
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Comamonadaceae; Comamonas.
1 (bases 1 to 7080)
Proident1,M.A., Mampel,J., MacSween,S., Cook,A.M. and Wyndham,R.C.
Comamonas testosteroni BR6020 possesses a single genetic locus for
extradiol cleavage of protocatechuate
Microbiology 147 (Pt 8), 2157-2167 (2001)
MEDLINE 21387526
PUBMED 11495993
REFERENCE 2 (bases 1 to 7080)
Proident1,M.A., Mampel,J., MacSween,S., Cook,A.M. and Wyndham,R.C.
Direct Submission
TITLE Submitted (13-SEP-2000) Ottawa-Carleton Institute of Biology,
Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S
2K9, Canada
FEATURES
source
1..7080
/organism="Comamonas testosteroni"
/mol_type="genomic DNA"
/strain="BR6020"
/db_xref="taxon:285"
141..1538
/note="putative transporter; member of aromatic acid:H+
symporter subclass of major facilitator superfamily"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdB"
/protein_id="AAK73575.1"
/db_xref="GI:14861219"
/translation="WTSQTSKQNTDPTVNVQTVLNSPFGQWGIFFLCFLIVLLD
GPTDAIGTIAISGLVHDQIGTIFETITGLGLGAMPNATITMNEYCPDSKRSFTNC
AAYVMGGAISAGLVHDQIGTIFETITGLGLGAMPNATITMNEYCPDSKRSFTNC
MFCGPICGAGGPTAANKIPIHFCWRSLLIIGGVVPLILAVLMIWLPESVRFVILKN
YSARVRKVMYIARAGTNRARFVLDNGEAAQAQQGPPQGLVFSOKILVFTLALW
VTYFMGLVIVYVGLVNMVPLFKAGVPASQAATAALFQLGCGVGAIFVGLLMDRWAN
LIATGYFTLSLGVAFIQGVLLGGVGLVAVSFIAGLLMNAQSSMQALAAEYPTEC
RASGVSMGLGTRGCGIAGSFLVLAELSRRLLEPLVFMVVAIPGVIAALALLVKRYA
GGAPRTSMGLGGAH"
1579..2607
/note="putative 4-oxalomesaconate hydratase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdB"
/protein_id="AAK73569.1"
/db_xref="GI:14861213"
/translation="MIIDVGHYTTAPALGAWRLQIAGLKDPKSTPSVADLKISDD
EICTETNQLRMKRGSDLTIFSPASFMAHHTIGDFTSSTWAAI CNELCFRVEL
PDHFI PAAML PQSPGVDPATCIPBLVKCEQGVNGVNLNPDPSGGHVSPLSDKS
WYPYIEKWEVDYIPAMIHVSTSCNSCHETGSHYINADTPAQMLCTSDLPKDFPTLX
FLIPLHGGANVPHWGRFGLAQEMKPLLEHLNNIYFDTCVYHQPGINLLTEVPT
KNILFASMEIGAVRGIDPQTGHYYDTRKYITEATQNLTADEKHAYEGNARVRFTLD
KALKAKGL"
repeat_region 2611..2634
/rpt_type=inverted
stem_loop 2641..2669
/evidence=not experimental
repeat_region complement(2687..2709)
/rpt_type=inverted
2741..3424
/note="putative 4-oxalocitramalate aldolase"
CDS
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdB"
/protein_id="AAK73570.1"
/db_xref="GI:14861214"
/translation="MYELGVVYRNIOQRADRAADGLAALGSATVHEAMGRVGLLKPYM
RPIYAGKQSVSTAVTLLQPGDNMMHVAEQIOGDIIVAATVAECSDGVFGDLLAT
SFQARGARALIIDAGVRDVKTLQEMDFPWSKAISSKGTIKATILGSVNIPIVCAGMLV
TPGDVIVADDDGVVVCVFAARAVEVLAQAQKRESFEGERAKLASGLVGLDMYKREPL
EKAGLKYYID"
3603..4520
/note="putative 2-pyrone-4,6-dicarboxylic acid hydrolase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdB"
/protein_id="AAK73571.1"
/db_xref="GI:14861215"
/translation="MSQFEKTPGLDWYANPSKQFKLPAGAVDAHCHVFGNGNEFFV
APERKYTPCDASKAQLYALRDHLGFARNVVVQATCGADNEMAVDACKSVGGKARGVA
TVKRSISDAELQQLHDAGVRGVRNFARLVDFDPKDELMEIAGRIAKLGHVVIYFE
AVDLPGLWDFFTALPTTVVDHMGKPDVTGVDSEEFALFLKFMREHQNVMSKVSCEP
RLSVTGPKALNGEQAYRDVVPFARVVEEFDRVLTGTDWPHNLKDMFDDGLLVD
PIPIAPTAELQQLKLLVDNPMRLYPEEV"
4602..5051
/note="alpha subunit of protocatechuate 4,5-dioxygenase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdB"
/protein_id="AAK73572.1"
/db_xref="GI:14861216"
/translation="MALEQSKYLDVPGTIFDAEQSKRGYWLNQFCMSLKAENRRFR
ANEAYLDENWATEEQKQAVLRLNWCMTGGNIYFLAKTGATDKSPQAGSMGTG
MTEBYRAMMMGGGRSABENRYVGEDGDAQAHQPPQGSAGNQNGEN"
5055..5924
/note="beta subunit of protocatechuate 4,5-dioxygenase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdB"
/protein_id="AAK73573.1"
/db_xref="GI:14861217"
/translation="WARTASVFTSHVPAIGANDMGKTOEYAWAPLEKGYDFSRQWM
KONPDVVLVNDHATAFSLDCIPTFAIGTAEEQPADEGMPGRPVKPVGVHPDLAS
HIAQSVIOQDFDLTIVNMDVHGLTVPLSLMGQEDPKTSWPCPVFAVNVVQYP
VPTGRCFNLGRALKAVESYDQDINVTGTMGSHQLOGARAGLINKEWMDNQFLDL
LVENPHGLAQMPHIDYVREAGSEGIELVWMLIARGMSDVDPAPLPKVAHRFYHVA
SNTAVGHLLLENQ"
6014..6973
/note="putative 2-hydroxy, 4-carboxymuconic semialdehyde
dehydrogenase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdC"
/protein_id="AAK73574.1"
/db_xref="GI:14861218"
/translation="MSKTIKVALAGAGAFGINKHLDGKNIIDGVEVSVLVGRFDQTK
VADKYIAHVATDLAELALPEVDVILCTPTMAEQAIACMKAGKHVQVEIPLADA
LKDAVEALOKOGLVAMVGHTRFPNPSHOWVHKKEIAGENIOQMDVQTYFFPRTN
MNAIQARSWTDHLLWHAHTVDLPAYQAGSPIVKANVQSPITKDIAGIAMSIOQL
KAANGAICTLSLNSFNNDPLGTFRTYIGDTYTLARYDDLYLHDEQLQNLAD"
7010..7041
/evidence=not experimental
stem_loop
ORIGIN
Query Match 4.0%; Score 20; DB 1; Length 7080;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



QY 330 GTTCAAGGAGCGGCGGTGC 349  
|||||  
Db 1028 GTTCAAGGAGCGGCGGTGC 1047  
MAU66560 7853 bp DNA linear BCT 08-NOV-1996  
Mycobacterium avium Embr (embr), EmbA (embra) and EmbB (embB) genes,  
complete cds.  
U66560 GI:1619916  
Mycobacterium avium  
Mycobacterium avium  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterinae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
avium complex (MAC).  
1 (bases 1 to 7853)  
Belanger, A.E., Besra, G.S., Ford, M.E., Mikusova, K., Belisle, J.T.,  
Brennan, P.J., and Inamine, J.M.  
The embAB genes of Mycobacterium avium encode an arabinosyl  
transferase involved in cell wall arabinan biosynthesis that is the  
target for the antimycobacterial drug ethambutol  
Proc. Natl. Acad. Sci. U.S.A. 93 (21), 11919-11924 (1996)  
97030297  
8976238  
REFERENCE  
2 (bases 1 to 7853)  
Belanger, A.E. and Inamine, J.M.  
Direct Submission  
Submitted (09-AUG-1996) Microbiology, Colorado State University,  
Center Street, Fort Collins, CO 80523-1677, USA  
LOCATION/Qualifiers  
1. 7853  
/organism="Mycobacterium avium"  
/mol\_type="genomic DNA"  
/strain="2151"  
/db\_xref="taxon:1764"  
complement (1. 1155)  
/gene="embR"  
complement (1. 1155)  
/gene="embR"  
/function="regulator"  
/note="ethambutol target"  
/codon\_start=1  
/transl\_table=11  
/product="embR"  
/protein\_id="AAC44546.1"  
/db\_xref="GI:1619917"  
/translation="MANAGLFGVLPSVTAAGQALPLGAPKQRAVLAMLLIHRNRP  
VSDALIDGVGAAPVPAARTSIQSYVSTLRRLRLSALPNPNGLIASVPPGYQLNVAD  
ACDLGRFRGHTAGVOAASGRFEASGHLAALREWRGPDLDLHNFVDAFANL  
LLERVAHTARAEAEACGRADGVIQELALVAQHPYREPLNAQLITAYVTERQSD  
ALGRYRKTALEAGLIDPGTFRALQORILROEPGPGOPGGAARAMLSTHKQST  
VRSEAADVNDPVVALRDKAGHYQLNGVTTRIGRLDNEIVLDDTEVSRHVAII  
DTGDFLITDMKSTNGVQRGRVRSQATLVLDGDIHRIQNSFEVFEIRPA"  
1156. 1332  
/gene="embAB"  
1156. 1332  
/gene="embAB"  
/note="divergent"  
1333. 4659  
/gene="embA"  
1333. 4659  
/gene="embA"  
/function="arabinosyl transferase"  
/note="ethambutol target"  
/codon\_start=1  
/transl\_table=11  
/product="embA"  
/protein\_id="AAC44547.1"  
/db\_xref="GI:1619918"  
/translation="MPHDGKQSRQSRPSRVAAGIAGLLCLAVPLLPVRQTATVL

WPQGTVDGHVSQITAPLVSGAPRALDISIPCAVATLPADGGLVWSTLPPGGMDAGKN  
GLFRANKDVVVAFRDTVAQAQRPAAAGASVHLHAWADAGAAGAEFYIGPAAGT  
LPAEKPGYGGFTDLKVPAGPLSARVDIDTRFTAPTVLKOIVAVLGLVAVLITAV  
ALAVLDRSRGGGTLINRSPIALWSRYRGGTHLANRRVGLATWADAVLATLILW  
HVVGATSDGYNLTIIARVAPKAGLVYRFGTTDAPDWTGLLSRLASVSTAGV  
WMRLPATLIGICWLLIISHWLRRLQRRGGLAAANRVAVFTAGVFAAWLPFNNGLR  
PEPLIALGLVLTWMLVERAILQRLAPAAVAVVALLTATLTAQGLIAVAALLTGARA  
VQAQIRRRASDGLLAPLAAALSLILVVFVSQTATVLESARIKYKVGFTIAY  
QWMLRYFTLVESNPDGSMARPAVLWMLCLFGLMVLILRRHVPVAGSGPWRLLIG  
TVAGLLLTFTKWAQVQFAGLAGALGALTAFACSRIGLHNRNLTLYVALLF  
VLWATSGINGMFYVGNVYDIOQVITASHVTSFELTSLITGLLAQWFRMDY  
AGHTEVDSRRNRVLAETLLVVATIMVGVESLTKGAFVRIYLTITKANAAARAS  
GLPTSCAMADDVLAEPDANAGMLPQLPGQTFGDPGLGCVNPGKPDGVDGDDLOS  
PVVTKPGLVNSDASPNKPNVAYSDSAGTAGGKGVNGVSHAALPGLDPAFTVMGS  
YGENSLAATASATSAWYLPRTDRLPVLVWSAAGAIWSYKEDGFTTQOSLKLQWVAP  
PGSTVPLAEVQPIDIGPOPAENLRFELAWAPEANVARIVAYDNLSEONFAFTP  
PRVYVETVQLIGSOTPMMDIATAANPCQRPSEHLGVAELPAYRIILPDKQTA  
SSNLWSSSAGGPFLLQALLRTSTIYLRGWDYRDWGSVEQYFRVLVPADPADDAI  
EQGVMTVHGWSRQPIRALP"  
4656. 7853  
/gene="embB"  
4656. 7853  
/gene="embB"  
/function="arabinosyl transferase"  
/note="ethambutol target"  
/codon\_start=1  
/transl\_table=11  
/product="embB"  
/protein\_id="AAC44548.1"  
/db\_xref="GI:1619919"  
/translation="MSVSTVGGDVVRVWVATAGLIGFVSVATPLLPVQTTATLN  
WPQGLNSVTAPLISLTIPVDLTATVPCSLVRDLPPGGVILSTGPKKDAALNALF  
VVAHGKRVDTDRNVIIASARDQVAGAGCSRIEIASTRAGTATFVGLDPAKPLG  
GGPDPNLRPOIVGVFTDLTGAPAPAGLKSATIDTRFTSTTTLKLAAMVTAIATV  
ALVALWLDOLGHERLIPANRQFTLADAVIFGVFLWHVIGANSDDGYTLGWA  
RVADAGMNSYRNFSGSPDPFGWYNLLALMTHVSDASLWMRLPDLFAGIWCWLL  
SRVLPRLGPAVAASRANWAGVLLTAMWPFNDGRLPEPILALGSLVITYLLERSM  
RYSRLTPAALAVITAAFTLGVTGLIAALVAGGPPILRLILRRHVRVGTWPLVAP  
MLAAGTILTVFADOTLATVLEATRIATIGSQAWYTNLRYYLILPTVDSLSR  
RFGFLITCLCTAVFTMLRRKIPGVARGPAWLMGVI FGTMTFLLPTKWHFG  
LPAVAGNAMAALTIVLSPAVLWNSRNEAFLALLFMALCEATTTGWVVSQYGP  
FNSMTPKIGGIVTFTVTFSMFAALYAIWLHFASRHEGRLARALTAAVPLAAGF  
MALVTASWVAGIVRQYTSNADNIRREFSGCLADDVIVPEPDSNVNPGGDI  
LPGPLGGQHPGVGSFNGVPEHTVAEAIRTPNQPDYDWDAPTKLSAFNGSTVP  
LPLGLDAARVLGAGSYTQAQOSRLTSAMYRLPAPDDGPDVVVTAAGIAGNSVLH  
HHTDGTVLEVGRPGGGDIVPAGRLVVDLYGEOPKAWNLRFARSMDPADTVAR  
VVAEDLSLTPEDMTATPPPVBMRSIQEVYSTQPVLMQWAVGLAFPCQPMHVG  
VTEIPKRIPTDTAKQDITDWDGNGSLGILGITDLLLRAHVMSTYLSHDWGRDWS  
LRRFETIADHAPQALDGLGTATRTGWSPPGPIRKP"

## ORIGIN

Query Match 4.0%; Score 20; DB 1; Length 7853;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 TCGCGCTGCCGCGACGCTG 198

Db 2339 TCGCGCTGCCGCGACGCTG 2358

## RESULT 35

AE005116

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AE005116 12097 bp DNA linear BCT 12-FEB-2001  
Halobacterium sp. NRC-1 section 147 of 170 of the complete genome.

AE005116 AE004437

AE005116.1 GI:10581743

Halobacterium sp. NRC-1

Halobacterium sp. NRC-1

Archaea; Euryarchaeota; Halobacteria; Halobacteriales;

Halobacteriaceae; Halobacterium.

1 (bases 1 to 12097)

Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M.,



Shukla, H.D., Lasky, S.R., Baliga, N., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A., Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H., Isenbarger, T.A., Peck, R.F., Pohlschroed, M., Spudich, J.L., Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ehardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and DasSarma, S.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
205044483  
11016950

## REFERENCE

## AUTHORS

2 (bases 1 to 12097)

Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A., Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H., Isenbarger, T.A., Peck, R.F., Pohlschroed, M., Spudich, J.L., Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ehardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and DasSarma, S.

Direct Submission  
Submitted (14-JUN-2000) Institute for Systems Biology, 4225  
Roosevelt Way NE, Seattle, WA 98105, USA

## FEATURES

## source

1..12097

/organism="Halobacterium sp. NRC-1"

/mol\_type="genomic DNA"

/strain="NRC-1"

/db\_xref="taxon:64091"

76..699

/gene="cysg"

/note="synonym: VNG2332G"

76..699

/gene="cysg"

/note="Cysg"

/codon\_start=1

/transl\_table=11

/product="uroporphyrin-III C-methyltransferase"

/protein\_id="AAG20438.1"

/db\_xref="GI:10581744"

/translation="MLAVDPTGAI PRGDAVVVVVSKTGVELAADS GMTPGGRLAAI

GPATADCREAGMTVDVPEEYTSAGLVAALGDIDGAEVVARSDHGS AVLTGDLAE

AGAVHEVTVLVATPGAGGSTETAAAGDLDAALEFTSLTVEHLAARERGISSAA

REGLDATVGAIGPTAETPAADHGISVDVPPDDAGFEALARATIEDA"

738..2165

/gene="VNG2333C"

738..2165

/gene="VNG2333C"

/note="conserved hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="Vng2333c"

/protein\_id="AAG20439.1"

/db\_xref="GI:10581745"

/translation="WDTGPPVPALAEAAACADALRDADSVLVASHIDADGLTSAGI

ASTALERAAPHEVFSKQLDDEEADTANAGHTVLFDFGSGQDALTHERAGEF

TFVIADHQPAENAAATYHLNPLEGVGDSGLGAGATYVLRALLGTDATDLAALA

VVAGVGDRTVDGELVGNANRVAADGVEAGTLDATGALYGTQTRPLRLLEYASEL

YIPGTDHQGAI RFLDLDVLDHGDGWRWTWASLDHDBERRTVWSALVKRAVTVGPS

AEVTLTVGTYYTLTDEEPTALDASEPATLLNATARVERADVGLAVCMGDRGDALER

ARSLREERQTLSEGWSWTVREGTVTVDLQWFAHDDRIRETIVGIVAGMALSGDETE

NDRPIVAAKODEPAKRVSRATARKVVDAGVDLAVVVREAAAAGGGGGGHEAAGA

TVPAENERAFLLDAAGSVLAARQGG"

complement (2174..6079)

/gene="VNG2334C"

complement (2174..6079)

/note="conserved hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="Vng2334c"

/protein\_id="AAG20440.1"

/db\_xref="GI:10581746"

/translation="WTLTDAAGASPVHVLYVNDDDDFAEALTRRKLQSAFAGTFTTVG

TAALARLDESADCVCYSLSLPTTGGIDLLERLHDHDDPPTLFTGRGSERTASE

ATAGVSDYLFHAGQKSFELGARVRRLTEAARKETVAEMSDRFRILERSITDAVY

AVDEMRVYNNETMADRIDRDRAVGVTLWEFFPSLVGTPLEEQYRTAMETGSEVS

SEQLGAPFDYVDVRAFPDDDDGLTVFSQETDEREQALALERSETILRHHDVVVV

DDAGVGFANAAQRTQSQAPLVGHHLPTIIEDRGSTADATSDVAIDSTLAEL

SDGTAGFYDNLPIDFDAGTDTTRTLDVRFPTFAADROVLVVGRODTGOSDVRRHL

EREDALEIQRVAETELSVDERLEALLAVGCOALGLEIGIVSHIHGEQYTVKAAT

SGVDIAPGEDFLATTICESVGDSDVCSFSDATADGRETHPAYEELGQSYVGAPLI

VDGRTYTNFSSPATRVADFAETEELLELLAEVLSAELSGRDAELQRQFLFER

IQETANGVFYFAPSTGSLWSGVRIRHDPGYPFDDAIDAHYDDRETTAAAV

DRALDEGPYDLRLIRPDRGVDRVWANGERSVSPQHGARTLRGVIQDITGRKGRER

EKHAAREYBALLTSGDAIFLLDVDTGDRPREFARLSPGYEAQTGLETAAVKGT

PREVGGQSEVAANRVCYDEGAPISYRELDVGPDAFWDTGLAPVVDGETVRI

VGIARVTDQVERSLSEATNQRLESLEATPLAWMEDPDGTVTRWNDEAESMGWT

REEVVEFNFPVDSQDRDEFDRHQRVLGGERRIGMERKRTKSGGDDVLLSAAAPV

GADGETTSLAVLEDITPKQLEAKRALQETARLTSAQSSAAVEAVEAADVLP

YDLTAVMEHDQSSASLVPAAATGATREAFDDIPTLQSGDSLAWDAFESAELHYDDLE

HQSLDVASGTPLESLGFLVPLGFGVIGVATAGTFSDTVDLFRILORIESAHDMA

TRETELQRNERLQDFASVAVHDLNPLSVATGFLDIADTGELEYLDRIESAHDMA

RLIDLLTLASGESITVDACVDLHNAVTTESNGYVDTABALCYRDDTLTSGDSRL

TOLFNFRNVAHEHGAAVTVTGOLGDRGFIYVADGDGIFEDERGERDVLDRHGVTS

AGTGFGLSIVEDIKAKHGWVVVTAGADGGARFARFETDP"

6285..7070

/gene="VNG2335H"

6285..7070

/gene="VNG2335H"

/note="hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="Vng2335h"

/protein\_id="AAG20441.1"

/db\_xref="GI:10581747"

/translation="WLYCSGGFGHGDYDGNRARGVPRGSGQASTAGPVRVILRS

QQRSGCTRAITRQDASIGLEPPPTWGLAPSRARWTRLRPTASHKDIPTIQNGTGT

RAVVRTDDGTRRRPSTPKAGSHPPAGVLCAGSGDHLQAGVQTTMSDMLTPFEFEQ

KYDYFPLETAYKRAFDMMGTVDKRLVHAIQQOVLSESEPFYEADSGFSEVLEPADP

LDRLAGVQVDADHARDVIDEHTDRICVHLRQFGIGGGQADEKA"

7096..7446

/gene="VNG2337C"

7096..7446

/gene="VNG2337C"

/note="conserved hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="Vng2337c"

/protein\_id="AAG20442.1"

/db\_xref="GI:10581748"

/translation="MSAESQDGSLEERVDFNFRNFQIQMHGGTAAVQNNINREDGV

VDLQLGACSGGIGSPMTIQALKSEMTKEIPDIDVNVAGTNGDSSGSGTSPFGEDGD

TEESAEDGQPQAPF"

complement (7474..11586)

/gene="pol2"

/note="synonym: VNG2338G"

complement (7474..11586)

/gene="pol2"

/note="Pol2"

/codon\_start=1

/transl\_table=11

/product="DNA polymerase type II large chain"

/protein\_id="AAG20443.1"

/db\_xref="GI:10581749"

/translation="MRPDETYFTELETGLDFAFVAEARERGSDPTNVEIPVAKD

MADRVENLIGDVAERVDRDLDGENREAEALVELVDFVDRGVGYDDTAGLEGAVR

TAVALLTREGVAAAPTEGIDRVENDNDGDTQVAVYVYAGPISAGDQAQLSVLEAV

ARAMLGIDAFPRDDEIERVAAEVDLYSETGLQYSPKDAETFTTTEHPCVMDGEAT

GNEFYDGRDLERIDTNSPRGMCVLVLAEGIALKAPKIQRYTRNLDEVAWPMQLDID

GIIGADDADETPDAGSDSDATDEGDASASTDAEPPRAASDFKRLRLIAGRPVFG

HEKNGGFLRYGRARNHGNATAGVHPATMHLDDFLATGTQIKTERPKAGAGIYVVD

SIEGTQVILANGDVRHINDPADLVRNGVAIILDVGYLVNIGYFVVENHLEPASY

APEWIQLDIAAGADVQALRDSPPYDLTAPTADQAMATDYDAFLHPAYTYLWHDID

VFQFALADADAHTDSDDDRGVLVDHTTTVRETLLEALLVTHHQGDDTIRVDDW  
 LFLARSLGVTSELDHEWETLSAAAEWENAVNEVAFPSVQERAPTEIGNMRGPE  
 KSESDLSFAVHTLPIFGAGSGQRDAADAAIAFADMSDTGPEIPRVGDRVCPSCDE  
 HYYEFCPCDGTDPHYECRCGAVATPEDESRVPCNGCRDLDNDVTTQVLDINDEY  
 HGALRAVGERENAFDOLGKVGILLSAEKTPEPMAGKVLRAKHDTAFKDGTVRYDMDT  
 LPTVTPAELDVTAQOFRELGVNODIHGDPLEHDDQLVELRVQDVVLSDGAADMLK  
 TADVDLLTOYGLDADFVLDLDRDLVGLVFGMAFPTSAAVGVRVAGFTSAGVYA  
 HYYFAKXRNCEPHTNWFDESGEWHDPLETLVEARLPDPTADEDDFGALVQAL  
 DGDVFPVTEGDEETLQRYEAVSRKAPADHLLAVETKRGRELTYTPDHSMERWTGCG  
 IERVDALTEAGELDPAFTPGDGETATSELRSGDGTGHPORRGCGSGSVRTIDGV  
 SVEPVRSVDHYSLVATNTLVANGLFTGCGDGEDCVMILMDGLINFPSYLPDK  
 RGRMDVMSGRIDPAEIDDEAHNIDIDREYFREFYATRELADPEDVADILITAE  
 SVTGTDEYTGFGHTATHNHLGSPISAVKTLGSMWMDKMDAOLELARKLRSVAETDV  
 AERYVHYHFLPDGLNLRAPSOETCLDCGKRYRMLSGDCRCGGRVNLTVHSGS  
 VNKMIDAMRVATEYDCREYTKORLIMRRLSVFDEDDTNKQSGISDFM"  
 complement (11586..12005)  
 /gene="VNG2339C"  
 /genes="VNG2339C"  
 /complement (11586..12005)  
 /genes="VNG2339C"  
 /note="conserved hypothetical protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Vng2339c"  
 /protein\_id="AAG20444.1"  
 /db\_xref="GI:10581750"  
 /translation="MARREVTGGREVFVRDPCSEWTELETAATEDITAAWFTGLG  
 AVQDAELWYDQDDFETAAEFPEPLEVACVGNIAITLDGEPFAHTATLSRRSGQAL  
 AGHLERATVAGELYVRELDADLVREHDAATDLDRL"

gene  
 CDS

Query Match 4.0%; Score 20; DB 1; Length 12097;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 GAGCGGGCGTGTGGACAC 356  
 |||||  
 Db 1299 GAGCGGGCGTGTGGACAC 1318  
 |||||

RESULT 36  
 AE004862/c  
 LOCUS  
 DEFINITION Pseudomonas aeruginosa PAO1, linear BCT 19-FEB-2003  
 Genome.  
 AE004862 AE004091  
 AE004862.1 GI:9950716  
 Pseudomonas aeruginosa PAO1  
 Pseudomonas aeruginosa PAO1  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 1 (bases 1 to 17589)  
 Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,  
 Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,  
 Lagrou,M.J., Garber,R.L., Goltry,L., Tolentino,E.,  
 Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,  
 Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,  
 Wong,G.K.-S., Wu,Z. and Paulsen,I.T.  
 Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
 opportunistic pathogen  
 Nature 406 (6799), 959-964 (2000)  
 20437337  
 10984043

JOURNAL  
 MEDLINE  
 PUBLISHED  
 REFERENCE

2 (bases 1 to 17589)  
 Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,  
 Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,  
 Lagrou,M.J., Garber,R.L., Goltry,L., Tolentino,E.,  
 Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,  
 Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,  
 Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,  
 Hancock,R.E.W., Lory,S. and Olson,M.V.  
 Direct Submission  
 Submitted (16-MAY-2000) Department of Medicine and Genetics,

TITLE  
 JOURNAL

University of Washington Genome Center, University Of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 3 (bases 1 to 17589)

Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)  
 Direct Submission  
 Submitted (04-FEB-2003) Department of Molecular Biology and  
 Biochemistry, Simon Fraser University, 8888 University Dr.,  
 Burnaby, British Columbia V5A 1S6, Canada

-----  
 This represents the February 3, 2003 version of the continually  
 updated, reviewed, Pseudomonas aeruginosa PAO1 genome annotation,  
 from PseudoCAP (see http://www.pseudomonas.com for latest updates  
 and links to alternate annotations). PseudoCAP is coordinated by  
 Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert  
 E.W. Hancock (University of British Columbia, Canada). We welcome  
 submission through www.pseudomonas.com of any proposed changes.

'Protein name confidence' is used to rate our confidence of the  
 accuracy of the protein name.  
 Class 1: Function experimentally demonstrated in P. aeruginosa.  
 Class 2: Function of highly similar gene experimentally  
 demonstrated in another organism (and gene context consistent  
 in terms of pathways its involved in, if known).  
 Class 3: Function proposed based on presence of conserved amino  
 acid motif, structural feature or limited sequence similarity  
 to an experimentally studied gene.  
 Class 4: Homologs of previously reported genes of unknown function,  
 or no similarity to any previously reported sequences.

## FEATURES

source

Location/Qualifiers

1..17589

/organism="Pseudomonas aeruginosa PAO1"

/mol\_type="genomic DNA"

/strain="PAO1"

/db\_xref="taxon:208964"

complement (75..569)

/gene="mreD"

/locus\_tag="PA4479"

complement (75..569)

/gene="mreD"

/locus\_tag="PA4479"

/note="Protein name confidence: Class 2 (high similarity  
 to functionally studied protein)"

/codon\_start=1

/transl\_table=11

/product="rod shape-determining protein MreD"

/protein\_id="AAG07867.1"

/db\_xref="GI:9950717"

/translation="MARATGGSNNVWVWISLLVGLVAVAPMTFTTEIGRPLMLML  
 IYVWLLPVRGVMTAWLLGLAQDVLVGNLGNALITGLITFTVLSLHQLRMFPA  
 WQCCVLVWVYGLAQLQLMLNALTGNRPPTLFLPALVSALLMPWVYALLQFVRLR  
 LNVK"

complement (569..1561)

/gene="mreC"

/locus\_tag="PA4480"

complement (569..1561)

/gene="mreC"

/locus\_tag="PA4480"

/note="Protein name confidence: Class 2 (high similarity  
 to functionally studied protein)"

/codon\_start=1

/transl\_table=11

/product="rod shape-determining protein MreC"

/protein\_id="AAG07868.1"

/db\_xref="GI:9950718"

/translation="MLAVLSVLSMVVDARFDYLEPVRSKLGNVLTFFVGLAEMPVRAM  
 EGVDRQSSRELSIAENRKAESILMORRVOKLAUTEQNVRLKELNSAALVDVVKV  
 LVSSELIGVDNPFQTRIMDKGENDGVFGQFVLDASGLMGVVEWMEYTVARVLLTLD  
 TTHSPIVQVNRGLRAVAGTGNPERLERLYVADTADIKEDGLVSSGLGQFFAGYP  
 VAVKEVHDGSGQPFVAVRVATPMKRSRYLLVFSRSTPEQRDAABAQAEADK  
 KAAGAQAPQAPAAAPSPATPAQAQAAPAPAPATQPAAPANGRR"  
 complement (1641..2678)

gene

CDs

/gene="mreB"  
/locus tag="PA4481"  
/note="Synonyms: envB, rodY"  
complement(1641..2678)  
/gene="mreB"  
/locus tag="PA4481"  
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein)"  
/codon\_start=1  
/transl\_table=11  
/product="rod shape-determining protein MreB"  
/protein\_id="AAG07869.1"  
/db\_xref="GI:9950719"  
translation="MFKRLGMPSSDLSIDGTANTLIYVRGIVLNEPSVVAIRSH  
GSKSVVAVGPEAKMLGRTGNIAIRPKMGDVLADESCEKMLQYFINKVHNSFL  
QPSRVLICVCKTQVRRALRESALGAGAREVFLIEEPMAAIGLGFVEARGSM  
VVDIGGGTTEIALISLNGVVAESVRVGGRDFDEAIVTYVRNNGSLGISTAEIRKQ  
EIGTAFPGDVRGVDNRVLAEGVPRSFITNSNEVLEALQESLATTIVQAVKSALEQS  
PELASDIAERGLVLTGGGALLRLDLKLAQETGLPVIVASEPLTCVARGGGRALEMM  
DRHSMDLSTR"  
2918..3208  
/gene="gatC"  
/locus tag="PA4482"  
2918..3208  
/gene="gatC"  
/locus tag="PA4482"  
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein)"  
/codon\_start=1  
/transl\_table=11  
/product="Glu-tRNA(Gln) amidotransferase subunit C"  
/protein\_id="AAG07870.1"  
/db\_xref="GI:9950720"  
translation="MALERSDVEKIAHLARLGLSEADLPRTETILNWLGLIDMQAV  
DTSGVEPLAHLEATQRLRPDAVTEDHRDAYQTIAPAVEGLYLVPKVES"  
3221..4675  
/gene="gatA"  
/locus tag="PA4483"  
3221..4675  
/gene="gatA"  
/locus tag="PA4483"  
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein)"  
/codon\_start=1  
/transl\_table=11  
/product="Glu-tRNA(Gln) amidotransferase subunit A"  
/protein\_id="AAG07871.1"  
/db\_xref="GI:9950721"  
translation="MLHOLTIAEIAIALADKQFSAEELTRTLGRIOLDPOLNSFIS  
ITDDLAIAQAAADERRANGALLGAPIAHKDLFCQGVRTSCGSKMLDNFVSPYD  
ATVVEKLTAGAVTLGKLNDEPFAMGSSNQSHYGAVKNPWSLDRVPFGSGGSAAV  
AARLPAAATGDTGSGIKOPAAITNLGTIKPTIGRVWGMTAYASSLDQGLFART  
EDCALMGVAGFDKDSVEQPYDDYLAALQKPLSLGRIGLPRFYFGAGLDSIAD  
AVLAVBELTIGATVKDISLPNMQHAI PAYVYIAPAEASNLSPFDGVRGYRCDAF  
ONLEDYKRSRAEGFSGVKNRIMVGTVALSAGYDVAVYLOAKIRRLIKNDVFAFA  
EVDVLTGPTNPANKIGEKNDPVSQYLEDIYVITITANLGLPGLSMPAGFVGLPVG  
VQLLAPYQEGRLNVAHQYQQVSDWHTRTAGF"  
4782..6227  
/gene="gatB"  
/locus tag="PA4484"  
4782..6227  
/gene="gatB"  
/locus tag="PA4484"  
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein)"  
/codon\_start=1  
/transl\_table=11  
/product="Glu-tRNA(Gln) amidotransferase subunit B"  
/protein\_id="AAG07872.1"  
/db\_xref="GI:9950722"  
translation="MQWETVIGLEIHAQLATQSKIPSGSSSTAFAAPNTQASLVDLAM  
PGTFLVNEAEVMAACLGIAIDRIDRONFAKNYFYPDLPGKYQTSQMDHPITVGK  
GHLDITLEDGTTKRICITITRAHLEEDAGKSLHEDFQCMGSDIDINRAGTPLEITVSEPD

gene

CDs

Query Match 4.0%; Score 20; DB 1; Length 17589;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 GGAGCGCGACGCGCTGG 277  
|||  
Db 10600 GGAGCGCGACGCGCTGG 10581

RESULT 37  
SRERYAB  
LOCUS  
DEFINITION  
ACCESSION M63677.1  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 20235)  
AUTHORS Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. and Katz, L.  
TITLE Modular organization of genes required for complex polyketide biosynthesis  
JOURNAL Science 252 (5006), 675-679 (1991)  
MEDLINE 91220065  
PUBMED 2024119  
COMMENT  
FEATURES  
source  
Location/Qualifiers  
1..20235  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1836"  
19..20235  
/gene="eryA"  
19..10722  
/gene="eryA"  
/note="eryA ORF2 encoding modules 3 & 4 for 6-deoxyerythronolide B formation6-deoxyerythronolide B formation; putative"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA26494.1"  
/db\_xref="GI:152694"  
/translation="MIDSEKVAEYLRRATLDLRAARQRIRESPIALVSMACRLPG  
GYNTFQVRLLEGGETLSGFPPTDRGWDLARLHPDPNPGTSTVVDGGLDDAGF  
DAEFTFGSPRAAAMDPOQLLETLSWELVNEAGIDPHSLGTATGVLGAKFVGE  
DTAAEDVEGYSVTGVAVASGRISYTMGLEGPISVDTACSSSILVALHVAESLRK

```

GSSMAVVGCAAVMATPGVVFDFSRQALAAADGRSKAFGAGDAGFGFSGEYTLVLLER
LSEARNGHEVLAVRGSALNQDGSNGDSAFSGPQARRVIRIQALSSCGLERFGDVDAV
EAGHTGTALQDPIEANVALLDTYGRDNDADRPLWGSVKSNIQHTQAAQVGLLKVVVL
ALRNGELPATLVEETPHVDSSGALLAGNPWRGRTERRARVSASFISGNTAH
VIVEAPERHEHRETTAHDGRVPLVVSARTTAALRAQAQIAELLERPADLAGVLG
LATRARHEHRAAVASTREEAVALREIAAGAATAADAVGVTEVDGRNVVPLFPQO
GSOWAGMAGELLSSSSVFAGKIPACDESAPMDKVSVDROAPGAPGLDRVDPVGP
VLFVNVSLAEIWRSYGVZPNAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRLRS
LSGEGMAAVALGEAAVRRLRPMQRLSVAANGPRSVVSGEPGALRAFSDCHAE
GIRVDIDVYASHSPQIRREBELLETTDITAPRPARVTFHSTVESRMSDFTELDAR
YWRNLRETRVADATRIAEVSDYDAIEVSPHFVVQVVEEAVEEADGAEDAVVGS
LHRDGDGLAFRLSMATRVSGVDIWDVALPGAAPFALPTYPQKRYWLOPAAAPA
ASDELAVRYSMTIEXPEGSNLDGMLVVTPLISPEWTEMLCEAINANGRLRCVDD
TASRTEMAQAQVATGGRGVLSLLSSDESACRGVPGVAGGLLTINQALRGACVDA
PVMLTQGAVRTPADDDGLPAQTTAHGFAQVAGLELPGRWGVVDLPESVDDAALRL
LVAVRGGRAEDHLAVRDRHGRVVRASLPSQSRSTPHGTVLVLTGAASPVGDQ
LVRLWADRGAERLVLAGCPGDDLLAAVEEAGASVVCADQAAALREALGDEPFTALV
HAGTLTNFGSISVAPPEFAETIAAKTALILVDELVRDRAVERVVCSSVAGIWGGA
GMAAYAGSAYLDALAEHHRARCSCTSVANTWPMALPGGAVDGYLRERGLRSLSDR
AMTWERVLAAGPSVAVADVDPVLSSEGPATPTALPAELAGRGQAEEPDSCPT
GEPALRGLSPDEQENLELAVANAEVIGHESAAEINVRRAFSELGLDLSNMAAL
RKLSTAGLRLPASLDFHTPTVTAQAHLRAELVGDQAQVAVVGADESEPIALV
GICRPFPGGIGSPGIRVLAEGANLTTFPADRGWDIGRLYHPDNPCTSTYVDGFG
FLTDAADFPGPGFIPREALAMPQORLMLETAWEAVERAGIDPDALRGDTGTVFGV
MNGOSYMOILLAGEAERVDGYQGLGNSASVLSGRIAYTFWEGEPALTVDIACSSSLGI
HLAQMALRGECSLAGGVTVMSDPVTYFVDFSTQGLASDGRCKAFSAPADGFALSE
GVAALVLEPLSRARANGHVIVIRGSANQDQASGLAAPNGPSQERVIRQALASG
VPAADVVVBAHGTGTGLGDIIEAGALITATYQDRDRPLRGLSVKXNIGHTQAAQAA
GVTKVLMHMGMLPSRLSHADELSPHIDWESGAVEVIREEPWMPAGERPRAGVSFG
VSGTNAHVIVEEAPAEQARTERRGPLFPVLSGRSEAAVAAQARALAEHLRDTPELGI
TDAANTLATGRAPDVRAAVLGDGRVCAELDALAEGRPADAVAPVTSAPRKPVLV
FPQCGAQWGMADLLESSVFASMSRCALSPHTDWKLLDVVRDGGDGPDEHVRDQ
VLPQVLFISIVSIAELMRHGVTPAAVVGHSQGEIAAAHVAGALSLEAAKVVVALRSQ
VLELEDQCGMVSQASRDELTVLARQDQSVAAVANGPSTVWAGTVAELDEFFAE
AEAREMPKRIAVRYASHSPSEVARIEDRLAAELGTTITAVRGSVPLSHSTVTGVDITSA
MDASVYWRNLRLPFAQVRLVGEQFDTFVEVSPHPLVLMAVEETAEHAGAEVTCV
PTLRGOSGPHFELNLLRAHVGVGADLRPAVAGGRPAPLPTVPPEHQRFPMPRHP
ADVSALGVRGAEHPLLAAVDPVGHGAVTGLRSTDEQPLAEHVGGRTLVPGSVL
VDLAAAGEVDHGVLPVLEQLRPLVLAGALLRMSVGPDESGRKTIIDVHAEDVA
DLADAQWSHATQAQVAAAGPRDTEQWPPEDAVRIPLDDHYDGLAEQGYEYGPSFO
ALRAAMKDDSVAEVSTAEADDEGYAFVPLLDVAOTLSLIGALGEPGGKLPFAWNT
VTLHAGATSVRVVATPAGDAMALRVTDPAHLVATVDSLVLRVSTGEKVEQEPGRGG
EGELHALDMWGLABGSGTRVVAADSDLDVLESREPEPDAVLVRYEPGDDPRAAA
RHGVLWAAALVRWLEQBELPGALVTATSGAVTSDSDSVPEPGRAAMGVLRCAOA
ESDPFVLLTDABPGMLPVPFNPQALRGDDVFPRLSPLAFSALITFAGTORLVP
GDGAIDSVAFEPAPDPOPLRAGEVRVDVATGVNFRDLVLLALGMPKADMGTEAAG
VVTAVGPDVDAFAPGDRVLGLFQGAFAPIAVTDHLLARVDPGWSADAAAVPTIAYTT
AHVALHDLAGLRAGOSVLIHAAGGVMAAVALARAGAEVLA TAGPKHGTALRALGL
DDEHIASSETGFKARKERTGRGVVDVNLSTLGLLDESADLLAEDGVFVEMGKTD
LRDAGPRGRYAPDLEAGDDRLGEILREVVGLLGAGLEDRLPVSAWELGSAALQ
HMSGRHVGLVLTQAPVDPDGTGLTGTGTGLRLRLARHLVTEHGVRLHLLVSRRG
ADAPGSDLERAEIEDLGASAEIAACDADTADALSDLLDGLPRUTGVVHAAGVLAADGL
VTSIDEPAVEQVLRAKYDAWNHLELTANTGLSFVLFSSAASVLAGPQGVYAAANE
SLNALALRRRGLPKALGMLGVAQSEMSTGLGDRIARTGVAALPTERALLFDISA
LNRGEEVFPPLSINRSALRAAEFVPEVLRGMVRAKLAAGQAAGNVPVLTLAGRSE
SDQVAGLAELVRSHAAVSGSADQQLPERKAFDGLPDSLAAVLELBNELGTANGVRL
PSTLVFHTPLTVAEHRDRKLFAPASPAVDLGRDLELEKALAEALSLEDGDDHVGQRL
ESLURMNSRADAPSTSAISEDASDDLSLDFMLDQRFGGGEDI"
19. .4470
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of module 3"
97. .1482
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of beta-ketoacyl ACP synthase
of module 3"
1693. .2670
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of acyltransferase domain of
module 3"
3406. .3921

```

```

/genes="eryA"
/notes="putative"
/function="approximate span of beta-ketoreductase domain
of module 3, possibly non-functional"
4171. .4428
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of acyl carrier domain of
module 3"
4471. .10722
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of module 4"
4471. .5847
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of beta-ketoacyl ACP synthase
domain of module 4"
6054. .7026
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of acyltransferase domain of
module 4"
7165. .9216
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of dehydratase and
enoylreductase domains of module 4"
9433. .9984
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of beta-ketoreductase domain
of module 4"
10225. .10483
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of acyl carrier domain of
module 4"
10723. .20235
/misc_feature
/genes="eryA"
/notes="ORF3 encoding modules 5 & 6 for
6-deoxyerythronolide B formation; putative"
/codon_start=1
/protein_id="AA026495.1"
/transl_table=11
/db_xref="GI:152695"
/translators="MSCDNMTTEKKRYLKRTVTLELDSVTLRREVEHRAEPIALV
GMARFPGVDSPESFWEFVSGGDAIAEAPADRGWEPDPDARLGGMLAAGDDAGF
FGISPREALAMPQORIMLEISWEALRAGHDPVSLRGSATGVTGTGYDGRPDE
APDEVLGVTGTASSVAGRVAYCLGLEGPAMTDTACSSGLTALHAMESLRDEC
GLALAGVTVMSSPGAFTEFRSOGGLAADGRCKFPKAAQDGFGLAEGAVLVLORLSA
ARREGRVLAIVLRGSANQAGNLTAPSPACQORVIRALENAGVRAGDGDVDEAH
GTGRLGDPLEVHALISTYGAERDPPDLMLGWSVNIHTQAAAGVAGWKAVLALR
HEAMPRTHTPEDEPSQIEMDLGAVSVVQARSAPAGERPRRAGSVSFGISGTNAHVIV
EEMPEDEPPAPDPSPGPVPLVLSGRDEQAMRAQAGRLADHLAREPNSLRTDGTFLAT
RGSABWHRVAVVGDRLDALAGLADRIADRTATGOARTRGVAMVFPQCAQWOG
MRDALLPESOFADSIFDCERALAPHVDMSITDLSGARPLDRVVDVVPALFAVWVL
AALNRSGVPAVVGHSQGEIAAAHVAGALTLEDAAKLVAVRSEVLRGLGGQGMAS
FGLTEQAAERIGHFAGALSIASVNGPRSVVAGSGPDLDELITAECAEGITARRIPV
DYASHSPQVLESLELTLAGISPVASDVALYSTTQOPIIDTAMDTAYMYANREQ
VRFODATRLQALAEAGFADFVEVSPHPLVTGTEATLSDALPADAGACVGTLRDRDGL
ADPHATLGEAYAGQVEVDWSPADARPVELPVYFQORVWLPPIPTGGRARDDDW
RYQVWTEAEWESASLAGRVLLVTPGVPSPSELSDAIRSGLEQCATVLTCDVSRSTI
GTALAEADTALSTVLSRDEAVDPSLDALALVQALGAAGVEAPLWLTNRVVOV
ADGELVDPAQMMGGVLGVGIEQGRWGLVDLVDADAASIRSLAALVADPRGEEQV
AIRADGKIKVARLVPAPARAARTSRSPVTLVTGTCGIGAHVWARLWESAEHLVLL
GRGADAPGASSELRELTATGVTIAACDVADRLEARLEAARAEGRVTVAVMHAA
GVSTSTPLDILTAEAFETIADVKVGTVNLDELCPDLDAFVLFSNAGVWSPGLASV
AAANAFDGLFARRRSESGAPVTSIANGIWAQNMAGDEGGEYLSRQGLRMDPRAVE
ELHITLDHGQTSVSVVMDRRRRFVELFTAAHRLPEIETAGARAEARQSESGPAAQR
VLTSTAEAREHRLAHLIRAEVAVLGHGDDAIDDRFADLGRFSDFTADPDRNLAA
VTVREAAATVVDHPTITRLADHYLERLVGAEEAQAPALVRFVFKDADDPDIAIVGMA
CRFPGGVNPGELWEFTVGRGDAVTEMTDTRGWDLDALFDPPDQRTGHTSGRGAFLD

```

ORIGIN

RBS	gene	CDS
-----	------	-----

```

46. .20268
/gene="eryA"
46. .10749
/gene="eryA"
/notes="ORF B"
/codon_start=11
/transl_table=11
/product="6-deoxyerythronolide B synthase II"
/protein_id="CAA44448.1"
/db_xref="GI:581651"
/db_xref="GOA:Q03132"
/db_xref="SWISS-PROT:Q03132"
/transl_table="MTDSEKVARYLRRATLDLRAARQRIRELESDDPIAIVSMACRLPG
GVNTPQRLWELLREGGETLSGFTDGRGDLARLHPDPDPNGTSYVDKGFGLDDAAGF
DAEFFGVSPREAAAMPQQRLLLETSWELVENAGIDPHSLRTGTATGVLGVAKRGYGE
D3AAEDVEGYSYTVGAPAVASGRISVYTMGLEGPSISVDTACSSILVALHILAVESLRK
GESSMAVGGAAWMTDPGVDFSRQRLAAGDGRKAFGAGADGFGSEGTVLLVLER
LSFARRNGEVLAVRGASALNQGDSNGLSAPSGAQRVIRQALESCGLEPGPDVAV
EAMGTGTALGDPTAEANALDITYGRDNDADRPLMLGYSKSNIGTQAAAGVTLGLLV
ALNGEPLATLHVEEPTPYDMSGGVALLAGNQPRWRGRTRTAAVSAFGICSTNAH
VITTEAPRERHRETTAHDGRPVLVVSARSTAAALRAQAQIAELLEPPDLACGVGLG
LVAARPHREAVASTREEAVRGUREIAAGATAADAVVEGVTEVDGRNVVFLFPGCG
GSOWAGNAGELLSSPVPFAGIKRACDESMAQMDKWSVDLVRQAPAGLDPRVDVQF
VLFAVWYSLAEKSVSGYVEFAAVVGHSSQGEIAAAHVAGALTLEDAKLIVGRSLRSL
LSGEGGMAAVALCEAAVRELRPWDELSVAAVNGPESVSVVSGEPGALRAFSEPCAAE
GIRVRLDDVYAGSHSQIERVREELLETGTGDIAPRPARTVTFHSTVESRSDGTETLDA
YWRNLRTDVFADATRLAEBSGYDAFIEVSPHPVAVVEEAVEEDADGADAADVVGGS
LHRDGGDLISALFASMTAHSVGYVDIRWDVALPGCAAPFALPTYPQKRYVMLQPAAPA
ASDPLA VRVSWTPTKPESGNLGDGMLVVTPLISPEWTEMLCEAINANGGRALCRCEV

```

```

TSASRTMAQVAQAGTGRGVLISLSDSACRPGVPAGVGLLTLVQALGDGAVDA
PWLTCGAVRTADDLAPACOTTAHGAQVAGLPGRWGVDPDPESVDDAALRL
LVAVLGGGAEHLAVRDLGRVVRASLPQSGSRWTPHGTVLVTCGASPVGDQ
LVWLADRGAERLVLGACPGDOLLAAREAGASAVVCAQDAAALBALGDEPVTIALY
HAGTLTNFSGISSEVAFEEAFTAAKTALLAVLDEVPDRAVEREVCSSVAGIYGA
GMAAYAGASYLDAEAHHRARGSTSVANTFWALPGHVDGGLRGLRGLSLSDR
AMTWERVLAAQSVAVADWDPVLSEGFPAATPTALFAELAGRGQAAEPDSGPT
GEAPALAGISPEQOENLLEIVANAFAEVLGHESAAENVRFAFSELGDLSDNAMAL
RXLSASTGLRLPASLVPDHTPTALAQHLBARLVGDQAAQAVRVVGADESERAIIV
GICRFPGGISPEQURVLAEGANTLTFPPADRGMDIIRLXHPDNDPFGTSYVDKGG
FLTDAAWTATGRAPVDRAAVLGDDRGVCAELDALAERGPSADAVPTSPAPKPVIL
MNGQSYMQLLAGAEVRYQGLGNSAVLSGRIATYFGEWGPALTYDTRACSSLVGI
HLAQAALRRGECISLAGGVVMSDDPYTFVDFQSGVGLASDGRCKAFSARADGFALSE
GMAVALLEPLSRARANGHOVLAVRANVQDGSANLAAAPNGFSORVIRQALAAAG
VPADVAVHAGHTGTELGDIIEAGALIAATYGODDRPLRGLSVKTNIGHTQAAAGAA
GVIVKVLAMRHGMLPRSLHADELSPHIDMESGAVEVLRSEVPWAGERPRAGVSSFG
VSGTNAHVIVEEAPAEQEAARTERGLPFLVLSGRSEAAVAAQALAEHLURDTPELGL
TDAQWTATGRAPVDRAAVLGDDRGVCAELDALAERGPSADAVPTSPAPKPVIL
FPQCAQWGMARDLLESSEVFAESMSRCALSPHTDWKLLDVVRDGGDPDHERVD
VLOPVLFSIVYSLAEIWRHAGVTAAVUGHSQGSIAAAHVAGALSLEAAKVVALARQ
VLRDLDFGSMVSGARDELETVLARWDCGRVAAVANGPCTSVVAGTAELEDFPAE
AEREMKPRIVARIASHSEVARIEDRLAEIGLITITAVRGSVPLHSTVTGEVIDTSA
MDASYRNRLRPLVFPQAVRGLVEQGFDTFVEVSPHPVILLMAVEETAEGHAEVTCV
PTLRREGSGHEFLNRLIRAHVHGVDLRLPAGVAGPABLPYTFPHQRFWRPHRP
ADVSALGVGAHEPLLLAAVDVPGHGAFTGRLTDEQPLAEHVVGRTLVPGSVL
VDLAAAGEVDGLPTELVLRPLVLGAGALIRMSVGAPDESGRITIDVHAADVA
DLADAOWSHATGTLAOGVAGPDTQMPEDAVRIPLDHVDGLAEQGVYGPSPQ
ALRANWKDSVVAEVSIAADEGYAHPVILLDAVATLSLIGALGEGGKLPKAWNT
VTLHAGTAVRVVATPAGDAAMALRVDPDPAHLVATVDSLVVRSSTGEKWEQEPREGG
EGELHALDWRLAEFGSTRVVAADSDLAVALRSGPEDEPDLVLRVEPEGDPRAA
RHGVLMAALVRLWELELPGATLVATISGAVTSDDDSVPEFGAAAMGVTRCAQA
ESDPFVLDTGATGEMLPVAPDNPQALRGDDVFPRLSPALPSALTLPAGTQLRVL
GDGALDSVATEAPDVPQPLRAGVRVDRATGVNFRDVLILALGMYPKQKMDMTEAAG
VTVAGVDVDAEAPGDVGLGFOGAFAPITWDRILARVDPDGSADADAAPVTAFTT
AYHALDLAGURCAQSVLIHAAAGGVMAVALLARRAGAEVATAGPAXKHTLALGL
DDEHIASRTEFGARKYRGGRGVVDVNLGELLDESADLLEAGDGVFVEMGKTD
LRDAGDFRGYAPDLPAGDGRDGLTEILREVWGLLGGALDRLEPVSAMELGSAPAAQ
HMSGRHVGLVLPQAPDPDGTVLITGGTGLRLRLHVLTVHGVHLLVSRRG
ADAPCSDELRAIEDIGASAFIACDTRADRALSALLDGLPRLTGVHAGVLAIDL
VTSIDEPAVEGLRAKVDAAWNLHETANTGLSPFVLFSSASVLAGGVYAAANE
SLNALALTRVGLPAKALGWLWAQASEMTSGDGLRIARTGTGVAALPTERALALFDSA
LRREGVVFPLISNRSALRAEFYEVLRVMVRAKIRAAQAAQAEANVVRDLRAGRE
SDQVAGLAEVLRSHAAAVSGYSADQLPERKAFKDLGDFDSIAAELNRLRGTATGVRL
PSTLVFDPHTPLAVAEHLRDLFAASPAVDIGDLRLEKALEALSAEDGHDHVDGQRL
BSLLRWNSRBDAPSTSAISDASDDLEFSLMQRFEGGEDLI"
139. .1416
/misc_feature
/gene="eryA"
/product="beta-ketoacyl synthase"
1471. .2763
/misc_feature
/gene="eryA"
/product="acyltransferase"
4261. .4455
/misc_feature
/gene="eryA"
/product="acylcarrier protein"
4513. .5787
/misc_feature
/gene="eryA"
/product="beta-ketoacyl synthase"
5839. .7116
/misc_feature
/gene="eryA"
/product="acyltransferase"
7117. .7563
/misc_feature
/gene="eryA"
/product="dehydratase"
8743. .9234
/misc_feature
/gene="eryA"
/product="enoylreductase"
9418. .10164
/misc_feature
/gene="eryA"
/product="ketoreductase"
10315. .10509
/misc_feature
/gene="eryA"
/product="acylcarrier protein"
10750. .20268

```

```

/gene="eryA"
/note="orf A"
/codon_start=1
/transl_table=11
/product="6-deoxyerythronolide B synthase III"
/protein_id="CAA44449.1"
/db_xref="GI:46979"
/db_xref="GOA:Q03133"
/db_xref="SWISS-PROT:Q03133"
/translation="MSGDMGTEELRRYLKRTVTLEDSVTARLEVEHRAEPIAIV
GMACRPGVDSPESWEFVSGGDALAEAPKRGWEPDPAIRLGLMLAAAGFDAGF
FGISREALAMDQQRIMLEISWEALERAGHDVSLRGSATGFTGVTGVDYGRPDE
APDEVLYGTGTASVAGSRVQAYCIGLEGFAMTVDTCSSGLTGLHLMAMESLRDESC
GLALAGVTVMSPFGAETEFRSQGLAADRCRKFKAADGFLGAEAGVNLQRUSA
AREGEPVLAVLAGASVNDQASNGLTAPSGAQQRVIRRALENAGVAGVDYVEAH
GTCTGLDPLEVHALLSTVGAERDPPDPLWIGSVKSNIGHTQAAAGVWKAVLALR
HGMPRTLHPDESQPLENDLGAVSVVSOARSAPAGERPRRAGVSSFGISGTHAHVIV
EEAPEADPEPPADSGPVLVLSGRDEQNRQAQRLADHLAEPRLNSLDTCGTLAT
RAGAMERAVVVGDRDEALAGRAVDRIADRTATGQGENSRVAVNVPFGQAQWQ
GMARDLRESQVPAISIRDCERALAHPVDMSLTDLGSRPLDRVDVQVAFVWVS
LAALMRSHQVEPAAVVHSGQETAAAHVAGALTLEDAKLVAVRSVRLRLGQGGMA
SFLGTEQAAERIGREFAGALSIAVNGPRSVVVVAGSGPLDELIAECEAEHAKARI
PVDYASHSPVESIRELLTELAGISVPSADVALYTTTGQPIDTATMAYWYANLR
EQVYQDATQLAEAGDAFVEVSPHPVLTVGIEATLDSALPADAGACVGTLRDSG
GLADPHTLGEAYAQGVVEVDFAFADARPELVPFQRYWLPITPGGRARDDDDD
RYQVWREAEWESASLAGRVLLVTPGVPSELSDAIRSGLEQSGATVLCDEVERSTI
GTALEADTALSTVGVAAPHGEAVDPSLDALVQALGAAGVAEALVWLRNNAVQV
ADEGLVDPQAQMVGLGRVVLVIGQRWGGLVDLVDAADAASIRSLAALADPRGEEQV
ALPAGDKVAVLAPAPARTHPLEPLAGTVLTGTTGGLAHARLARLSAGSHVL
LGRGADPAGSAREELTALGTGVTIACDVAARLEAVLAAEAAREGTVSAVWH
AAGVSTPTDDETEAEFTIADVKVGTNLDLDELCDLDAFVLFSNAGVWGSFGLA
SYAAHLITLDHGTSVSVMDRRRFTAEARHPLEFDEIAGARASARQSEEGPDRA
VEELHITLDHGTSVSVMDRRRFTAEARHPLEFDEIAGARASARQSEEGPDRA
QRLAALICDGREREHLAIRAEAAVVAHLGHDDAAIDDRAFRDLGFSMTAVDLNR
LAATVGREAAVTVEDHPTITRLADHYLERLVGAEEAQAPALVREVPKQADPILAI
VGNACRPGGVHNPGELEWETVGGSDAVTQWETDRGMDLALPDQPHGCTVSRHG
AFDGAADFADAFGFSREALAMDPOORVLETTWLEFENAGIDPHSVRGSDGVLEL
GAAYQGVQDAVVPDESEGVLTTGNSAVSVAYVVLLEGPAVTALACSSLEVAL
HSACQGLDSDGLAVAGGVSVAGPEVTFEFSQGLAVDGRCKAFSAEADGFLPE
GVAVQVQLRSLDGLPAEGGRVGVVAGSAINQDGTNGLAAPSAQVQVIRKAWARA
GITCADVAVVEAGHTGRLGDPVEASALLATYKSGRSGSVLIGSVKSNIGHAQAAA
GVAGVIVKVLGNLGLVPPMLCRGERSPLIWSGGGVLEAVSPWPAPAGVVRAGV
SAFVGSVFLPILALASALATGRAHLPRAALAGDEHQRLQRAVAGVAAEPAGTGTGA
SAGVSVFVFCQAGQWEGMARGLLSPVFAESIACCDAVLSEVAGVASEVLEQRPDA
PSLRVDVQVPLFVSMVSLARLWAGCSPSIVIGHISQGEIAAAVAVVLSGDEVR
VVALRAKALPAGKGGWVSLAAPGERARALIAPWEDRIISVAANVSPSVVSGDPEA
LAELVACEDGVRKATLVDVYASHSRHVEIRETILANDLGLSARAAIPLYSTLHG
ERRDMGPRYWDNLRSQVRPDEAVSAQSPDGHATFVEMSPHPVLTAAVOEIAADAVAI
GSLHRDTAEELHIAELARAHVGVAVDMRNVPFAAPFVALPNYPFPEQVWIAPEVSD
QLADSVYRVDWRPLATTVDILEGGLVHGSAPESLTSAVEKAGGVPPVVASADRALAA
ALREVPGEVAGVLSVHTGAANALAHOSLGRAGVRAPLMLVTSRAVAGSESPVDEQ
AMWGLGVWGLETPERWGLVDLPASPGDGEAFVACLDGDEHDOVATRDHARYG
RLVRAPIGTRSSWEAGTALVTGGTGLKLEELRGQRPRVTVVHTAGVPSERPLH
AAELEALALGPKTITITACDADRQKLEELRGQRPRVTVVHTAGVPSERPLH
VAALIGHSPDVGQDQPFTELGFDSLTVGLNRLOQATGLALPATVLEFHTVRL
ADHIIQQDSDGTTPAREASSALRDGTROAGVSRVSYLLDLAGLSDPREHFDGSGFS
LDLVMDADPGGEVTVICAGGKFPVYVAGSAGALMAYALATELLDRGHPRPQTEGFLP
SSMAAAVQAVADVIRTOGDKFPVYVAGSAGALMAYALATELLDRGHPRPQTEGFLP
PPHQDAMNMLELTATLFDRETVRMDDTLTALGAYDRLTGQWRPRETGLPLLS
AGEPWGPPDSDSWKPTWTFEHTVAVPGDHTWQEHADATARHIDAWLGGGNS"
10867. .12111
/misc_feature
/gene="eryA"

```

Query Match 4.0%; Score 20; DB 1; Length 20444;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 124 TTCGACTCGTGGCGCGGT 143  
 |||||



```

Db      10381  TTCGACTCGTGGCCGGGT 10400

RESULT 41
E38021/c
LOCUS   E38021          31422 bp      DNA      linear      PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38021
VERSION   E38021.1 GI:18626910
KEYWORDS  JP 2000245457-A/2.
SOURCE    Streptomyces avermitilis
ORGANISM  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 31422)
AUTHORS   Omura,S. and Ikeda,H.
TITLE     Avermectin aglycon synthase gene
JOURNAL   Patent: JP 2000245457-A 2 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT   OS Streptomyces avermitilis
            PN JP 2000245457-A/2
            PD 12-SEP-2000
            PF 24-FEB-1999 JP 1999046961
            PR
            PI SATOSHI OMURA,HARUO IKEDA
            PC C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH Key
            FT CDS
            FT CDS Location/Qualifiers
            1..31422
            /organism="Streptomyces avermitilis"
            /mol_type="genomic DNA"
            /db_xref="taxon:33903"

ORIGIN
Query Match 4.0%; Score 20; DB 6; Length 31422;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 CCTCGCCGACCGGCACCG 385
|||||
Db 27210 CCTCGCCGACCGGCACCG 27191

RESULT 42
BD097650/c
LOCUS   BD097650          31422 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION A method for procuding avermectin derivatives.
ACCESSION BD097650
VERSION   BD097650.1 GI:22643224
KEYWORDS  WO 0162939-A/2.
SOURCE    Streptomyces avermitilis
ORGANISM  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 31422)
AUTHORS   Endo,H., Yamaguchi,H., Kanda,Y., Hashimoto,S., Omura,S. and
            Ikeda,H.
TITLE     A method for procuding avermectin derivatives
JOURNAL   Patent: WO 0162939-A 2 30-AUG-2001;
            KYOWA HAKKO KOGYO CO LTD,THE KITASATO INSTITUTE,HIROFUMI ENDO,
            HIROYUKI YAMAGUCHI,YUTAKA KANDA,SHINICHI HASHIMOTO,SATOSHI OMURA,
            HARUO IKEDA
COMMENT   OS Streptomyces avermitilis
            PN WO 0162939-A/2
            PD 30-AUG-2001
            PF 23-FEB-2001 WO 2001JP001381
            PR 24-FEB-2000 JP 00P 47405
            PI HIROFUMI ENDO,HIROYUKI YAMAGUCHI,YUTAKA KANDA,SHINICHI PI
            HASHIMOTO,

```

```

PI SATOSHI OMURA,HARUO IKEDA
PC C12N15/54,C12N15/53,C12N9/10,C12N9/14,C12N1/21,C07C327/30, PC
C12P17/08
CC A method for procuding avermectin derivatives FH Key
FT CDS
FT CDS Location/Qualifiers
1..31422
/organism="Streptomyces avermitilis"
/mol_type="genomic DNA"
/db_xref="taxon:33903"

FEATURES
source
1..31422
Query Match 4.0%; Score 20; DB 6; Length 31422;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 CCTCGCCGACCGGCACCG 385
|||||
Db 27210 CCTCGCCGACCGGCACCG 27191

RESULT 43
MSGY222/c
LOCUS   MSGY222          41156 bp      DNA      linear      BCT 03-DEC-1996
DEFINITION Mycobacterium tuberculosis sequence from clone Y222.
ACCESSION AD000010
VERSION   AD000010.1 GI:1702968
KEYWORDS
ORGANISM  Mycobacterium tuberculosis
            Mycobacterium tuberculosis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE 1 (bases 1 to 41156)
AUTHORS   Du,L.
TITLE     Direct Submission
JOURNAL   Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
            Beaver Street, Waltham, MA, USA, 02154 du@ctic.com
COMMENT   GSDB:S:1004711
FEATURES
source
1..41156
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
/clone="Y222"

ORIGIN
Query Match 4.0%; Score 20; DB 1; Length 41156;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 TCGGCCCGCCGCGGCGTTC 111
|||||
Db 14176 TCGGCCCGCCGCGGCGTTC 14157

RESULT 44
SFU78289
LOCUS   SFU78289          43280 bp      DNA      linear      BCT 13-AUG-1997
DEFINITION Streptomyces fradiae tyactone synthase, starter module and modules
            1-7, (tylg) gene, complete cds.
ACCESSION U78289
VERSION   U78289.1 GI:2317859
KEYWORDS
SOURCE    Streptomyces fradiae
ORGANISM  Streptomyces fradiae
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 43280)
AUTHORS   DeHoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
TITLE     Sequence of Streptomyces fradiae tyactone synthase gene tylg

```



JOURNAL  
REFERENCE

2 (bases 1 to 43280)  
DeHoff,B.S., Sutton,K.L. and Rostock,P.R. Jr.  
Direct Submission  
Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate  
Center, Indianapolis, IN 46285, USA  
tylg.

COMMENT

FEATURES

Location/Qualifiers

1..43280  
/organism="Streptomyces fradiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1906"  
1..43280  
/gene="tylg"  
816..14234  
/note="ORF1; putative"  
/codon\_start=1  
/transl\_table=11  
/product="tylactone synthase starter module and modules 1  
& 2"

source

gene

CDS

/protein\_id="AAB66504.1"  
/db\_xref="GI:2317860"  
/translation="MSSALRRVQSNCGYGLDLSNTAQTQDQEDVDGPDSTHGGE  
IAVVGNSCRIPGAAGVEEFWEILSRGMEPTQDDGTWRAALDHDAGFDAGFFGMNAR  
OAAATDPOHRLMLELWEALDAGIVPGDLTGTDTGVFAGVADDVAVLTRRSVAGS  
GYTAGLHRLAANRLSHFLGRLPSLVVDSQASIVAVOLACESLRGETSLVAG  
GVNLLTSESTWERNKGLSPDGRCHTFDARANGVYRGEGGAVVLKPLDALADGD  
RVYCVIKGGAVNNDDGSGALSTPDREAQAVLRQAYRAGVSTGAVRYVLEHGTGTRA  
GDVEAAALGAVLAGADSGRSTPLAVSGYKTNVGHLEAGAGIVGLIKATLVKRGEL  
VPSLSTPNPDLPLDRLRVTQTERQWNEEDDRPRVAGVSGFGMTGNHLVIRKGA  
PAAAGSGGAGSGGAGSIVSGVVPVSGRSRVVREAAAGRLAEVAGGVGLA  
DVATWADRSFCYRAVLARGAEALGRALAGDPPDAGVGTGVLDCGVVGGIAP  
GGAGAGGAGAGGAGGGGTVLPDQGTQVCMGAGLIGSSSEVFPAASRECRALSV  
HVGWDLLEVVAGGAGRLVDPQVPTVMVNSLARYVQWAVDVAVNVHSGSEIRAA  
TVAGALSLEDAAVVALRALGLIGRLAGRGMAAVALPAGVEAGLAKWPGVGAAN  
RSPRVSCVAGEQPEYVQCAEGVQARLIPVDYASHSRHVEDLKEGRFVLSGRIP  
VLHAIQTAEDADRSVHNGTLRRODSSPHRLTSTAEEAHGATLTWDPALPPGHL  
TTLPTFFNNHWDLTIDGGGDATQKESGLTRRLRGLPSSQKQLDPLDVCUR  
HTAVLGLDTAAEVDPLSKQGIQSMGTVELNRLLTGTGLATPILVIDRTPTRA  
LAQFLHTLDDSPSGSVLAPAKQSFQAQPIAVVGMGCRFPGGSGPEALMRLVFG  
VDVSPPPGDRGVGGLVDPKPGVAGSVYRGGLFLHDAEFDAEFPGISPREAVAM  
DQORLLLTSWAELBRAGIDPHSLHGRSVTVAGVMPQVGPRLAEGAGSGDGLYT  
GTSGSVVSGVAVTLGLEGPATVDTACSSSLVALHIAVQALRGECMDALAGVTVM  
AGPMVFERRQRLADGRKAFADGADGTANAGAGVILVERLSARLGHPLVAV  
VSAVNDQASNGLTAPSGPSQERVIRQALGNARLTADVDDVVEAHGTRGLDPIE  
AQALLGTYGRDGRFPVWLGSLKSNIGHAQAAGVAGVTKWVLMRYGLPRTLHVD  
EPRSHVDWAGVWLLTEAREWPGVDRPRRAAVSAFVSGTNAHLILEAPDPAESA  
TTPVRSVSSAAVLIDARSQVFPVVSGRSVRVREAGRLAEVVEAGVGLADAVT  
MAGSRGVAVLVARGAEALGRALAGDPPDAGVGTGAVVDVPGTSGGGGVLVF  
PGQQTQVWGAGLIGSSSEVFPAASRECRALSVHVVNDLLEVVSGGAGLERVDVOR  
VTWAVMSLARYQWGVDAVVSQSGEATAAVVAGALSLEDAVVAVALRAGLIGR  
YLAGRGMAAVALPAGVEAGLAKWPGVVAANGPASTVVSDDRRAVAGVAVCDAG  
GVQARLIPVDYASHSRHVEDLKEGRFVLSGRIPRVSCVAGEQPEYVQCAE  
YWFNLNRVFSVAVGGLLEGRHRRFIEVSAHPLVHAIQTAEDADRSVHNGTLR  
RODSSPHRLTSTAEEAHGATLTWDPALPPGHLTTLPTFFNNHWDLTIDPPTPAT  
TTQSTPDARVRYVTKALTSSVPRPHSIGELLVAPPTDGLLGLTTLVLSERGS  
VARLEVPFGARRAEVALLKFSMESAGEENTVSLGLGVLESTDAVRTSIALLOQVSD  
IGVFAARVMALTRRAVAVPGETPDQAGLWFGFVRAALDELPMWGLLIDPLDAEEL  
TRTPTSQPTPPTPQTPNRRALALEAAVLAGRDGEDQVAVRASGIYGRVSRRAA  
AGAAWSPQSVLITGMCAGLGRRLARLAAEAGRLVLTSSRRGPPEAPGAELAEELR  
GHGCEVHAACDVAERDALAALVTAVPNVAVFTAGIIDDVAIDTSLSPESVYRGAK  
VCGRELLHQLDRLKGLDAPVLFSRVGTWGNAGGAVAAANRALDALAERRAAGLP  
ATSVANGLWGGGMAAGAEESLSLGRWMDPPDAVALLGAMGRNDVCVTVDDVW  
BRFAPATNAIFRPLFDFTVPEAREALTAAAGTSATPDGAPELARLRLSMNLTERLKL  
VELVTEAARFHPNTDAIGAEERFPKSGFSLTSLRLNRNAGTGLKLPATVIFD  
HPSPTALARLLDRLTGAGAPADAPDELPAVADDPPVTVGMACRFPGGAGPEA  
LKLWTEREDVIGAPTRDGLDSDVDPPEGVAGTVYRGGLFHDAAEFDAEFGLI  
SPREAVAMPQORLLTSTWAEIRAGIDPHSLHGRSVTVVGLTHQRYASRLHAPE  
IAGVYLLTGKAGSVVSGRISYTLGLEGPSLSDITACSSSLVALHNAQAALRGECMDA  
LAGVTVMAAPGLFVFEFGORGLADGRKAFADGADGTANAGAGVILVERLSDDAR  
LGHPVLAVCGSVANNQDGSNGLTAPSGPSQERVIRQALGNARLTADVDDVVEAHGTC

TELDGPIEAQALLGTYGRDRAECPCVWLGSLKSNIGHAQAAGVAGVTKWVLMRYGM  
LPRTHVDEFSRHVDSAGGVRLTEAREWPGVDRPRRAAVSAFVSGTNAHLILEAP  
EALALEATDAPEAPADVDTVTEALEAPADVTEAREGAKAPGSPPEAQFVGVVP  
VVSGRSRVVREAAAGRLAEVAGVGLADVAVTMAGSRFGYRAVVLARGAEALAG  
RLRALLAGDPPDAGVVTGAVVDPETSGGGGVLVFPQGTQVWGMAGLIGSSSEVFAA  
SMRECARALSVHVVNDLLEVVSGGAGLERVDVQVPTWAVMSLARYVQWGVDAVVA  
VGSQGEITAAATVAGALSLEDAAVVVALRAGLIGRYLAGRGMAAVALPAGVEAGLIA  
KWPGQVAAVNGPASTVVSDDRRAVAGVCOAGVQARLIPVDYASHSRHVEDLKEG  
ELERVLSGIRPRSPRVVSCVAGEQPEYVQCAEGVQARLIPVDYASHSRHVEDLKEG  
HRRFTEVSAHPVLVHAIQTAEDADRSVHNGTLRRODSSPHRLTSTAEEAHGATLT  
TWDPALPPGHLTTLPTFFNNHWDLTIDPPTPATTTQSTPDARVRYVTKALTSSVPR  
DSGERVFRAGLAGSEHDLTDHVSQVTLIPGTAFVFEFALHAGAAATCGCRLSELSVEA  
PLVLPAAAGVVRVQMSRAADESGRRKVAIHSAPAAVHSAAGSGDSAGVWTRHGEGTL  
VPDPEPTPDADNARAWPAGERVPEAEIYRFAGLYGEAFAGAFAGVWQPDALL  
AEVLLPDRASTAGTGFPGFALLDAAALQVWLAGLLEVEDPAVLLPFAVQVGSLYATG  
AGALRVLTAKAGDGVSLQADTSGAAVSLGALVNRPLARRKLDVLLGLTDAGERSLY  
RVEMQRLIPAGPPRSWAVLGPADDELACTPGIGDQPDGPTALYPEVRLARKALAAV  
APRPEAVLVLSGAGATPESVRQTTERRCLTALQDWLDAEELVDTPLIVLTSGAAVAV  
PQEEIGDLACAGVWGLVRSARSEHPGFALVDTDHPDDRTALPLAIRAVLDGAGQLS  
LLRSGRAAAGADELAELAGLEAEPFAACDAADREALARLAEVADPPLTGVTHA  
AGVLDGDTDALTPERIGTVMRPKADAALNLHGLTSTPLSVFAVSGAAGILGRPCQ  
ANYAANTFEDDALAQRRAHGLPAYSLSLWGLWGGA"GMTGHLSTGLRMRRSIGIAPM  
THQGLALFDRAALASADPFLVPMRLDLAALVREARHGPDAVPGFLLGLLPARAVP  
RQAAAPVGGAPAGGEGTAERLAGEBEARLRELVRVRAEVSGVLGSGPDVAVP  
GRPPKDLGFDLSLTAVELNRNLGAATGLRLTALVDFRPTSQVAAEYLAELAGPRDGG  
DTAAAFEGLEALAAVAGALAEEDLARDVLRRLTLEAALTPQGRNPSAPAPASDL  
DRELSANDDDLLFAFTEEQL"  
942..3929  
/gene="tylg"  
/function="encodes starter module"  
942..2156  
/gene="tylg"  
/function="encodes beta-ketoacyl ACP synthase KS-Q of  
starter module"  
2571..3957  
/gene="tylg"  
/function="encodes acyltransferase domain of starter  
module"  
3675..3929  
/gene="tylg"  
/function="encodes acyl carrier protein domain of starter  
module"  
3993..8471  
/gene="tylg"  
/function="encodes module 1"  
3993..5264  
/gene="tylg"  
/function="encodes beta-ketoacyl ACP synthase of module 1"  
5631..6617  
/gene="tylg"  
/function="encodes acyltransferase domain of module 1"  
7410..7949  
/gene="tylg"  
/function="encodes beta-ketoreductase domain of module 1"  
8220..8471  
/gene="tylg"  
/function="encodes acyl carrier protein domain of module  
1"  
8541..13970  
/gene="tylg"  
/function="encodes module 2"  
8541..9812  
/gene="tylg"  
/function="encodes beta-ketoacyl ACP synthase of module 2"  
10260..11246  
/gene="tylg"  
/function="encodes acyltransferase domain of module 2"  
11319..11876  
/gene="tylg"  
/function="encodes dehydratase domain of module 2"  
12861..13415  
/gene="tylg"



```

/protein_id="CAD43449.1"
/db_xref="GI:27528335"
/db_xref="GOA:Q8GBX6"
/db_xref="SPTREMBL:Q8GBX6"
/translation="MSVTNDLTTEYELRLTQELHSETRLRATEERRHEPIAIVGLG
VLRPGIHDRDITLWTELEGRDAIPIPASRNADATYDPDAVGSKTVDAAMLDR
LDFDADFGIIPREAKYVDORHLLLETSMQLEADAGIVPDSIRSKTGCVFGTAS
VLRQSDRDAEAYAFMGMISSFAAGRLAFTLGLQGPALS IDTACSSSLHLACQ
SIQFQECDDLAVGVQVMSPEVFFVLSRTRALASGRSKTFSANAGDVGREGVQVL
AVERLDRAKGPPIIAVMSNHGTSGLTVPNGPQQVLRALDLDKALVPAD
VDFVECHGTGTSIDPIEYNALAAVYGEGRPKDPFLPGALKNTI GHELFASGLAVA
KMWASHTLTPATLSPINPLVDWALPVRVDAARPMWTRDDGAPRAGVTAIGL
SGNAAVIVEAATPEPTPDAPALPAVPLVLLSGDALKARQAARLAHLAAGRPD
ARLDVIAAATRTTHEDRAA VVAADRDELGCALDARGEAGPQSVASATPAGRV
VFFPQCGSQWGMARALLASSVVRDEITACERALLPHVMSIGVAGLSDEDAKIVA
GRVVQVPLFAVMAALWRSIGVTPDAVVGHSQETIAAATVAGALSLEDAKIVA
LVRDLAAGTAGMAAELGARDTEARLAPFGDAITAIAINSPRATLVAGDITADA
DGLRLDAAHFNRISVRPEADTRALHDDGHRFFVEVSPHVPVIALALDITLSIGD
AAVVSRLRDEGLERFLSAGELHARGHAI DWGAPFAPIGARRVTLPTVAFORFW
AARRSREQSTVGAWRYRVEWRPVYSAASRGDVAGTVLVVAPAGVSDLASALAATE
RGADVPLALEASDI TAHDASIRDRDAIGRPAIRGVITLPTDTPADRPAPVAGL
ARSLALVQALAEIAEPIWIVTSGVAIGRSDRIESPQVQAMWGIRGVVSLQEPERW
GGLVPLPRAHADGALDATALERLVAALGSEGEDELALRPTGLFARLARAPLGDAPA
QRTPEPRTVLVTCGTGALGHARVRLARHGAHLVLAASRRGAAPGAELAEITELG
ARYTIACDVADRAAIEBELFAALPAEPLTAI FHTAGVIDDGLVSLTPERLETVLHG
KADAARTLHBLTKRDLAAVFLPSSAAGVFGNAGQVAAANAFLDAAEERALGLR
ATSVAMGAGAGLFTDVAQGLRGFGFAEMEPIEIAVAALASALDRDTHVTIASIDW
ARFAPYASARQPLVADLPDARRALDQPSATSEHDELTALVATRALSGEARHLV
SLVLAQTAAVGHADRLSDLPQRFGLDLSLMAVELRLLOKATGKLPATITFDR
PSRHLGALVDLALPLCGAAPTSDARPAARADASDEPIAIVGVLLPLPGVDVGI
FRLEGREDVAPIPASRWDVGA VEDPDPEAKGTYVRHAAMLDRVLDFAFGFIS
REARVDPQRHLLETIAWQALEAATIVPASIRUSTGCVFVAGASDVAVILQSNBDE
AYAMWGASFAAGRLAFTLGLQGPALS IDTACSSSLVHLHACQSRKQECDDLAA
GVQVMAPEVFFVLSRTRALPDGRSKTFSANADGFGREGVVVLALELRALDRANHG
PDLAVRGVTAHNDGASGIPATFNGSSQKVLRAALDDARLAPTDVDFVCHGTGSL
GLTEVQALAAVYGEGRAPERLRLGAVKNTI GHELFASGLAGVKTIVASIRHGTLP
TIHNPENPHDIDALPVRVDAAPERSNADSPREAGVSAGLSCTNAHVIVEEAP
EPDAPKARGALPLSGALPVLVLSASDAALRAQAALRDQLARTPDALDVAAALAT
TRSQFDHRAALASDHGALVAALLEGVAAPAGTFAKGAADKLALFTGQGAQRAA
MGRGLYDAFVFRDALDVAHSLDRDLRDLVLPAPKSELSLDRDTEFTQPALF
ALEVALPRLVEAMGVTPDVLVGHSGVGLAAHAGVLSLSDACTVLAARARLMQALPA
RDGAMTVTHAEVLAALRGDRABETIAINAPSTTVAGDVDAVLAARVAARFARGR
KATLRYVSHA FSHHMDPMDADA FRVAEGLTFHPRIPIVSNVTGLAADAEIRSPDYW
VRHVSAVRADGINTLEADGVSSFFELGPHGVLSALGESALANAQREAVFVPAIRDG
RADVDALTAUSSLHVGRHVDAWFAFFGPDQTVLPTAFQDRFWLDRPAPAGTRD
EAPAPAEPAEFWRAVDAGDVGALGATINAGSEDLALATLPLALSARWRDERSL
VDALRYVWKPLTPTAARDVAGTWLLVTLAGAEADALARELERALTARGAEVITMPTA
PAEADRVLA VRVOGALADAGELRGIVSFAALDETTLATHTALPAGLALTLRLVQALG
DMGLI EAPLWLVTRGAVSGRDLASAAQSNMTGIVGVVLEHPPERMGGVLDVADGDT
IDARALDRVTLASRSGEDQALRPTGIFARRLVRAPLGDAPARTYTAGTSIYGT
GTGALGHARVFLAGRGAHLVLSRRGEDAPGAGARAELEALGARVTLAACDVDR
ASVAAALRLDAQDPVRAVHVAGVQAALASTDLDGVQAVVAAKVRGALHILHELL
GDRELDALVLPASGAGVWGSCQQAAGNAFLDALAEVRAAGLTATSTIAGWAGG
GMLADADADAEEVRLKRGVLPMA PALAIAALACALDHGETAVTVASVDMARFAPFA
SARPRLLDIP EAQRALEA PSAPAGADGASADRALVEALRPLSEADRARHLVVA
ETAIVLGHADASADPHKGFDFGLDLSLVAVELLRLKLRPTGVLELPATVTFDHPSPHR
LATVREALPDADAPADATATSSSAHAADIEPIAIVGLGRLPFGVEDREGIWSF
LEQQRDVGPTPKDRWDADAIYDPDPEAKGTYVRHAAMLDRVLDFAAFFGISPREA
KQVDPQRHLLETIAWQALEEATIVPASIKDTSTGVFVIGLSDYALMQSABEETDAYA
AMGTLSSFAAGRLAFTLGLQGPALS IDTACSSSLVHLHACQSRKQECDDLAGGVQ
VMTLAEPPVLLSRTRALADGRSKTFSADAGYGRGEGVVVLALERLERLDAEARGRVVL
AVVGTAVNHGASGITAPNGTSQOKVILRAALGDAGLAPADVDECHGTGALGDP
IEVQALAAVYGEGRAPERLRLGAVKNTI GHELFASGLAGVKTIVALEHGLPATIH
SRPNRPHVDMLPVRVDAAPERSNADSPREAGVSAGLSCTNAHVIVEEAPRAR
RGAATDPDPAPVLPVLSAKSERALRAQERLRAHVTGPGDALVDVAASLATTRSFH
EHRAAI VARDI GALEDALAGLEGRS PAGTVVSRITGKLLALFTTQCGQQAAMGRGL
YGA SFVFRDADFACSYLDREHPLRDVVPAPRGSSELAALIDETLTQTSIFALEVA
LFLRVVWGVTPHVLGHSTIGELVAHVAGVLSLEDACTLVAAARLMHALPRDDAM
VAVHFACEVHLDALDAGLDGRASIAAVNAPSSTVAVAGDREAIVAAVFEALGRKT
SRLRVSHA FSHHMDPMDAEFRVAGVFPFPRIPIVSNVTGARASNDILCSADYW
RHARGTVRFDGLEALRADGVSTFLELGHFAVLCAHGETHLESEAVAFPLALDRGRDD

```

Query Match 4.0%; Score 20; DB 1; Length 4800;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 187 CCCGCGACGCTGATCTTCGA 206

Db 38478 CCCGCGACGCTGATCTTCGA 38497

Search completed: June 16, 2004, 17:31:13

Job time : 2084.33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:31:50 ; Search time 246 Seconds  
(without alignments)  
8651.818 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543  
Perfect score: 501  
Sequence: 1 gcggtcccgacgcgcgacga.....ccctcgggcagacgagctga 501

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 15

Total number of hits satisfying chosen parameters: 2587

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq 29Jan04:\*

- 1: geneseqn1980a:\*
- 2: geneseqn1990a:\*
- 3: geneseqn2000a:\*
- 4: geneseqn2001a:\*
- 5: geneseqn2001bs:\*
- 6: geneseqn2002a:\*
- 7: geneseqn2003a:\*
- 8: geneseqn2003bs:\*
- 9: geneseqn2003cs:\*
- 10: geneseqn2004a:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	501	100.0	53789	2	AAV21187	AAV21187 Amycolato
2	31	6.2	9222	7	AAU61171	AAU61171 Actinosyn
3	31	6.2	82746	7	AAU61224	AAU61224 Actinosyn
4	25	5.0	44377	2	AAT80414	AAT80414 Platenoli
5	25	5.0	44377	2	AAT78508	AAT78508 Platenoli
6	23	4.6	748	4	AAU89129	AAU89129 Polyketid
7	23	4.6	753	4	AAU89143	AAU89143 Polyketid
8	23	4.6	6378	7	ACF39397	ACF39397 Mycobacte
9	23	4.6	31422	3	AAU92302	AAU92302 S. avermi
10	23	4.6	31422	3	AAU79278	AAU79278 Streptomy
11	23	4.6	110000	4	AAI99682	AAI99682 Continuation (19 o
12	23	4.6	110000	4	AAI99683	AAI99683 Continuation (19 o
13	21	4.2	47981	4	AAU30757	AAU30757 Micromono
14	20	4.0	4551	7	ACA42620	ACA42620 Prokaryot
15	20	4.0	29879	2	AAQ46806	AAQ46806 eryA regi
16	20	4.0	31422	3	AAU92302	AAU92302 S. avermi
17	20	4.0	31422	3	AAU79278	AAU79278 Streptomy
18	20	4.0	43280	2	AAU80413	AAU80413 Ty lactone
19	19	3.8	177	4	AAU17049	AAU17049 Human bre
20	19	3.8	177	4	AAU08770	AAU08770 Human bre
21	19	3.8	375	2	AAT73432	AAT73432 Human pre
22	19	3.8	375	2	ACA63335	ACA63335 DNA encod
23	19	3.8	375	3	AAZ35707	AAZ35707 Human uro

24	19	3.8	452	2	AAV99120	AAV99120 DNA methy
25	19	3.8	953	6	ABS63429	ABS63429 DNA encod
26	19	3.8	1049	6	ABS63428	ABS63428 DNA encod
27	19	3.8	1368	2	AAQ82825	AAQ82825 cDNA enco
28	19	3.8	1698	2	AAU86613	AAU86613 cDNA enco
29	19	3.8	1893	4	AAU54382	AAU54382 Pseudomon
30	19	3.8	1893	7	ACA42802	ACA42802 Prokaryot
31	19	3.8	2040	6	AAU85313	AAU85313 Human cyt
32	19	3.8	2448	9	ADD29815	ADD29815 Human tum
33	19	3.8	2478	2	AAU86614	AAU86614 cDNA enco
34	19	3.8	3081	6	ABQ88107	ABQ88107 Human oet
35	19	3.8	5092	4	AAU59458	AAU59458 Human pol
36	19	3.8	5434	6	ABL62902	ABL62902 Breast ca
37	19	3.8	5434	6	ABL63111	ABL63111 Breast ca
38	19	3.8	5434	6	ABL64407	ABL64407 Stomach c
39	19	3.8	5434	7	ABT33842	ABT33842 DNMT1 cDN
40	19	3.8	5794	5	AAU78583	AAU78583 DNA encod
41	19	3.8	19440	2	AAU99129	AAU99129 DNA methy
42	19	3.8	113193	7	AAU54645	AAU54645 Streptomy
43	19	3.8	128139	6	AAU64291	AAU64291 RRV genom
44	19	3.8	133719	3	AAU64754	AAU64754 Macaca mu
45	18	3.6	276	7	ACD93990	ACD93990 Human col

## ALIGNMENTS

## RESULT 1

AAV21187  
ID AAV21187 standard; DNA; 53789 BP.

XX AAV21187;

XX 24-JUL-1998 (first entry)

XX Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.

XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;

XX polyketide synthase; actinomycete; ansamycin; ds.

XX Amycolatopsis mediterranei.

XX	Key	Location/Qualifiers
XX	CDS	1825..15543
FT		/*tag= a
FT		/label= ORF_A
FT		/product= "Polyketide synthase"
FT		15550..30759
FT		/*tag= b
FT		/label= ORF_B
FT		/product= "polyketide synthase"
FT		30895..36060
FT		/*tag= C
FT		/label= ORF_C
FT		/product= "Polyketide synthase"
FT		36259..41325
FT		/*tag= d
FT		/label= ORF_D
FT		/product= "Polyketide synthase"
FT		41373..51614
FT		/*tag= e
FT		/label= ORF_E
FT		/product= "polyketide synthase"
FT		51713..5293
FT		/*tag= f
FT		/label= ORF_F
FT		/product= "polyketide synthase"

XX WO9807868-A1.

XX 26-FEB-1998.

XX 18-AUG-1997; 97WO-EP004495.

XX 20-AUG-1996; 96EP-00810551.  
XX (NOVS ) NOVARTIS AG.  
XX Schupp T, .Toupet C, Engel N;  
XX WPI; 1998-169172/15.  
XX P-PSDB; AAW52845, AAW52846, AAW52847, AAW52848, AAW52849, AAW52850.  
XX  
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used to  
XX produce rifamycin and rifamycin analogues.  
XX  
XX Claim 4; Page 53-102; 205pp; English.  
XX  
XX The present sequence represents a Amycolatopsis mediterranei rifamycin  
XX synthesis gene cluster DNA fragment from the present invention. The DNA  
XX fragment comprises a DNA region involved directly or indirectly in the  
XX gene cluster responsible for rifamycin synthesis, including the adjacent  
XX DNA regions to the right and left which, by reason of their function in  
XX connection with rifamycin biosynthesis, qualify as constituents of this  
XX rifamycin gene cluster, and functional fragments, derivatives or  
XX constituents of these. The Amycolatopsis mediterranei rifamycin synthesis  
XX gene cluster DNA fragment can be used for producing rifamycin, rifamycin  
XX analogues or precursors. It can also be used for inactivating or  
XX modifying genes involved in ansamycin or rifamycin biosynthesis. The DNA  
XX can be used for constructing mutant actinomycetes strains from which the  
XX natural rifamycin or ansamycin biosynthesis gene cluster has been partly  
XX or completely deleted. The DNA fragment can be used for assembling a  
XX library of polyketide synthases, which can be used for assembling a  
XX library of polyketides. A hybridisation probe of the invention can be  
XX used for identifying DNA fragments involved in the biosynthesis of  
XX ansamycins  
XX  
SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 U; 0 Other;  
Query Match 100.0%; Score 501; DB 2; Length 53789;  
Best Local Similarity 100.0%; Pred. No. 1.7e-218;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCGTCTCCGAGCGGAGCAAGCGATCTGTGTAAGCTGTTCGGGGGACGCTTCG 60  
DB 15043 GCGGTCTCCGAGCGGAGCAAGCGATCTGTGTAAGCTGTTCGGGGGACGCTTCG 15102  
QY 61 ACGETGTCTGCGCCACAGCGGCGCGAAGGCATCGCCCGCCAGCGCTTCAGAGATC 120  
DB 15103 ACGGTGTCTGCGCCACAGCGGCGCGAAGGCATCGCCCGCCAGCGCTTCAGAGATC 15162  
QY 121 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCCACCGGGCTG 180  
DB 15163 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCCACCGGGCTG 15222  
QY 181 CGGTGCGCGGAGCGCTGATCTTCACTACCCACCGGAGGCGTGTGGGTACCTG 240  
DB 15223 CGGTGCGCGGAGCGCTGATCTTCACTACCCACCGGAGGCGTGTGGGTACCTG 15282  
QY 241 CGCGTCGAACCTCTCGGAGGCGCGACGACGCGCTCGGAGGCGGAGACACCTCCGG 300  
DB 15283 CGCGTCGAACCTCTCGGAGGCGCGACGACGCGCTCGGAGGCGGAGACACCTCCGG 15342  
QY 301 CGAGTCTCTCGCGCGCGTGTCCGTCGCGGTTCAAGAGGCGCGGCGTGTGACACGCTG 360  
DB 15343 CGAGTCTCTCGCGCGCGTGTCCGTCGCGGTTCAAGAGGCGCGGCGTGTGACACGCTG 15402  
QY 361 CTCGGCTCTCGCGACACCGGCGCGGACCGGCGGACCGGCGGACCGGCGGCGGCG 420  
DB 15403 CTCGGCTCTCGCGACACCGGCGCGGACCGGCGGACCGGCGGACCGGCGGCGGCG 15462  
QY 421 CCGGCGCGCGGAGCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 480  
DB 15463 CCGGCGCGCGGAGCGACGACGACGACGACGACGACGACGACGACGACGACGAC 15522  
QY 481 GGCCTCGGCGGAGCGAGCTGA 501

DB 15523 GCCCTCGGCGAGCGAGCTGA 15543  
RESULT 2  
AAL61171  
ID AAL61171 standard; DNA; 9222 BP.  
XX  
XX AAL61171;  
XX  
XX 22-SEP-2003 (first entry)  
XX  
XX Actinosynnema pretiosum polyketide synthase (PKS) gene #2.  
XX  
XX Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;  
XX gene; ds.  
XX  
XX Actinosynnema pretiosum.  
XX  
XX WO2003045312-A2.  
XX  
XX 05-JUN-2003.  
XX  
XX 21-NOV-2002; 2002WO-US037547.  
XX  
XX 21-NOV-2001; 2001US-0332158P.  
XX  
XX (UNIW ) UNIV WASHINGTON.  
XX  
XX Floss HG, Yu T, Leistner E;  
XX  
XX WPI; 2003-493374/46.  
XX  
XX Novel maytansinoid produced by bacterial host cell transformed with  
XX expression vector comprising open reading frame from ansamitocin gene  
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX  
XX Disclosure; Page 53-58; 160pp; English.  
XX  
XX The invention relates to maytansinoid produced by bacterial host cell  
XX transformed with expression vector comprising open reading frame from  
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
XX useful as a potent antitumor agent. The present sequence is A. pretiosum  
XX ansamitocin gene cluster I polyketide synthase (PKS) gene  
XX  
XX Sequence 9222 BP; 784 A; 3437 C; 3858 G; 1143 T; 0 U; 0 Other;  
Query Match 6.2%; Score 31; DB 7; Length 9222;  
Best Local Similarity 100.0%; Pred. No. 0.00017;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 168 GGCACACCGGCTGGGCTGCGGCTGCGGCGGCGGCTG 198  
DB 5130 GGCACACCGGCTGGGCTGCGGCTGCGGCGGCGGCTG 5160  
RESULT 3  
AAL61224/c  
ID AAL61224 standard; DNA; 82746 BP.  
XX  
XX AAL61224;  
XX  
XX 22-SEP-2003 (first entry)  
XX  
XX Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.  
XX  
XX Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX  
XX Actinosynnema pretiosum.  
XX  
XX WO2003045312-A2.  
XX  
XX 05-JUN-2003.





```
FT FT CDS 31329..36071
FT FT /*tag= d
FT FT /note= "ORF4 encodes protein shown in AAW23719"
FT FT 36155..41830
FT FT /*tag= e
FT FT /note= "ORF5 encodes protein shown in AAW23720"
XX XX
XX XX EP791656-A2.
XX XX
XX XX 27-AUG-1997.
XX XX
XX XX 19-FEB-1997; 97EP-00301066.
XX XX
XX XX 22-FEB-1996; 96US-0012050P.
XX XX
XX XX (ELIL ) LILLY & CO ELI.
XX XX
XX XX Burgett SG, Kuhstoss SA, Rao RN, Richardson MA, Rosteck PR;
XX XX
XX XX WPI; 1997-418047/39.
XX XX
XX XX P-PSDB; AAW23716, AAW23717, AAW23718, AAW23719, AAW23720.
XX XX
XX XX DNA encoding Streptomyces ambofaciens platenolide synthase domain - for
PT PT production of spiramycin-related polyketide antibiotics.
XX XX
XX XX Claim 9; Page 8-33; 81pp; English.
XX XX
XX XX This sequence represents the platenolide synthase gene cluster of the
CC CC invention. This sequence is referred to as the smg gene, and was
CC CC isolated from Streptomyces ambofaciens. This sequence encodes the multi-
CC CC functional proteins which direct the synthesis of the polyketide
CC CC platenolide. Platenolide is the basic building block of the macrolide
CC CC antibiotic spiramycin. The DNA can be used to produce compounds
CC CC exhibiting antibiotic activity based on the platenolide structure.
CC CC including specifically the macrolide antibiotic spiramycin and spiramycin
CC CC analogues and derivatives. Modifications of the platenolide synthase DNA
CC CC sequence can be made so as to change the number and type of carboxylic
CC CC acids incorporated into the growing polyketide chain and to change the
CC CC kind of post-condensation processing that is conducted
XX XX
XX XX Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;
SQ
Query Match 5.0%; Score 25; DB 2; Length 44377;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 119 TCGGCTTCGACTCGTGGCCGCGGT 143
Db 35710 TCGGCTTCGACTCGTGGCCGCGGT 35734

RESULT 6
AA89129/c
ID AA89129 standard; DNA; 748 BP.
XX XX
XX XX
XX XX
XX XX 19-MAR-2001 (first entry)
XX XX
XX XX Polyketide synthase derived sequence 34-183R1A1.T7 DNA.
XX XX
XX XX Polyketide synthase; PKS; cosmid 34-183; FK-520; antibiotic;
XX XX immunosuppressive; neuroimmunophilin; ds.
XX XX
XX XX Streptomyces hygroscopicus.
XX XX
XX XX Location/Qualifiers
FH FT 1..747
FT FT /*tag= a
FT FT /partial
XX XX
XX XX US6150513-A.
XX XX
```

---

```
PD 21-NOV-2000.
XX XX
XX XX 16-SEP-1998; 98US-00154083.
XX XX
XX XX 16-SEP-1998; 98US-00154083.
XX XX
XX XX (KOSA-) KOSAN BIOSCIENCES INC.
XX XX
XX XX Wu K;
XX XX
XX XX WPI; 2001-049104/06.
XX XX P-PSDB; AAB19896.
XX XX
XX XX Novel nucleic acid encoding polyketide synthase useful for producing
PT PT recombinant polyketide synthase for making novel polyketides for use as
PT PT antibiotic, immunosuppressant or as neuroimmunophilin.
XX XX
XX XX Example; Col 20-21; 30pp; English.
XX XX
XX XX The present sequence is that of a polyketide synthase (PKS)-derived
CC CC coding sequence designated 34-183R1A1.T7. The DNA was isolated from
CC CC cosmid 34-183 of an FK-520-producing strain of Streptomyces hygroscopicus
CC CC following restriction digestion. It does not comprise a full-length
CC CC coding sequence. The 249-amino acid translated sequence is given in
CC CC AAB19896. An isolated nucleic acid comprising a coding sequence for a PKS
CC CC enzyme identical to, or isolated from, a PKS coding sequence contained
CC CC within a cosmid selected from cosmids 34-183, 34-122 and 34-126 is
CC CC claimed. Such nucleic acids, when used alone or in combination with other
CC CC PKS domain coding sequences, are useful in the construction of
CC CC recombinant vectors that encode PKS enzymes which can be used to make
CC CC novel polyketides having applications in medicine, agriculture and animal
CC CC health. The polyketides may have e.g. antibiotic, immunosuppressive or
CC CC neuroimmunophilin activities
XX XX
XX XX Sequence 748 BP; 114 A; 250 C; 277 G; 107 T; 0 U; 0 Other;
SQ
Query Match 4.6%; Score 23; DB 4; Length 748;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 169 GCCACCGGCTGGCGTCCCGC 191
Db 366 GCCACCGGCTGGCGTCCCGC 344

RESULT 7
AA89143/c
ID AA89143 standard; DNA; 753 BP.
XX XX
XX XX
XX XX
XX XX 19-MAR-2001 (first entry)
XX XX
XX XX Polyketide synthase derived sequence 34-183R1A.2 DNA.
XX XX
XX XX Polyketide synthase; PKS; cosmid 34-183; FK-520; antibiotic;
XX XX immunosuppressive; neuroimmunophilin; ds.
XX XX
XX XX Streptomyces hygroscopicus.
XX XX
XX XX US6150513-A.
XX XX
XX XX 21-NOV-2000.
XX XX
XX XX 16-SEP-1998; 98US-00154083.
XX XX
XX XX 16-SEP-1998; 98US-00154083.
XX XX (KOSA-) KOSAN BIOSCIENCES INC.
XX XX
XX XX Wu K;
XX XX
XX XX WPI; 2001-049104/06.
XX XX
```



XX Novel nucleic acid encoding polyketide synthase useful for producing  
PT recombinant polyketide synthase for making novel polyketides for use as  
PT antibiotic, immunosuppressant or as neuroimmunophilin.  
XX Example; Col 27; 30pp; English.  
XX  
CC The present sequence is that of a DNA sequence designated 34-183RIA.2,  
CC which was isolated from cosmid 34-183 of an FX-520-producing strain of  
CC Streptomyces hygroscopicus following restriction digestion. An isolated  
CC nucleic acid comprising a coding sequence for a polyketide synthase (PKS)  
CC enzyme identical to, or isolated from, a PKS coding sequence contained  
CC within a cosmid selected from cosmids 34-183, 34-122 and 34-126 is  
CC claimed. Such nucleic acids, when used alone or in combination with other  
CC PKS domain coding sequences, are useful in the construction of  
CC recombinant vectors that encode PKS enzymes which can be used to make  
CC novel polyketides having applications in medicine, agriculture and animal  
CC health. The polyketides may have e.g. antibiotic, immunosuppressive or  
CC neuroimmunophilin activities  
XX  
SQ Sequence 753 BP; 115 A; 253 C; 277 G; 108 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 4; Length 753;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 169 GCCACCGGCTGCGGCTGCCGCG 191  
Db 371 GCCACCGGCTGCGGCTGCCGCG 349  
RESULT 8  
ACF39397  
ID ACF39397 standard; DNA; 6378 BP.  
XX AC ACF39397;  
XX 26-SEP-2003 (first entry)  
XX Mycobacterium tuberculosis mycobacterial antigen DNA SEQ ID NO:172.  
XX  
XX Mycobacterium tuberculosis; mycobacterial; antigen; infection; vaccine;  
XX tuberculostatic; mycobacterial peptide; mycobacterial infection; gene;  
XX ds.  
XX Mycobacterium tuberculosis.  
XX WO2003033530-A2.  
XX  
XX 24-APR-2003.  
XX  
XX 14-OCT-2002; 2002WO-GB004647.  
XX  
XX 12-OCT-2001; 2001GB-00024593.  
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX James B, Bacon J, March P;  
XX  
XX WPI; 2003-393501/37.  
XX P-PSDB; ABM15903.  
XX  
XX New isolated mycobacterial peptide encoded by a gene that is induced or  
XX up-regulated under high oxygen tension, useful for diagnosing, treating  
XX or preventing a mycobacterial infection.  
XX  
XX Claim 18; Page 302-306; 392pp; English.  
XX  
XX The present invention describes an isolated mycobacterial peptide (I), or  
XX its fragment, variant or derivative encoded by a gene whose expression is  
XX induced or up-regulated during culture of a mycobacterium under  
XX continuous culture conditions of a dissolved oxygen tension of at least  
XX 30% air saturation measured at 37 plus degrees Celsius when compared with

CC a dissolved oxygen tension of up to 10% air saturation measured at 37  
CC plus degrees Celsius. (I) has tuberculostatic activity and can be used in  
CC vaccines. The mycobacterial peptide (I) or its fragment, variant or  
CC derivative, inhibitor, antibody, attenuated mycobacterium, attenuated  
CC microbial carrier, DNA sequence, DNA plasmid, RNA sequence, or RNA vector  
CC from the present invention can be used for manufacturing a medicament for  
CC treating or preventing a mycobacterial infection. The peptide or its  
CC fragment, variant or derivative, the antibody, or a polynucleotide probe  
CC comprising at least 8 nucleotides, where the probe binds to at least a  
CC part of the gene, is useful for manufacturing a diagnostic reagent for  
CC identifying a mycobacterial infection. The present sequence encodes a  
CC Mycobacterium tuberculosis mycobacterial antigen, which is used in the  
CC exemplification of the present invention  
XX  
SQ Sequence 6378 BP; 815 A; 1685 C; 2614 G; 1264 T; 0 U; 0 Other;

Query Match 4.6%; Score 23; DB 7; Length 6378;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ACGCTGATCTTCGACTACCCAC 215  
Db 6094 ACGCTGATCTTCGACTACCCAC 6116

RESULT 9  
AAA92302  
ID AAA92302 standard; DNA; 31422 BP.  
XX  
AC AAA92302;  
XX  
DT 10-JAN-2001 (first entry)  
XX  
DE S. avermitilis avermectin aglycon synthase DNA aveALL SEQ ID NO:2.  
XX  
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;  
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;  
KW agrochemical; ds.  
XX  
OS Streptomyces avermitilis.

XX Key Location/Qualifiers  
FH CDS 1. 14646  
FT /tag= a  
FT /note= "avermectin aglycon synthase protein"  
FT 14824. 31422  
FT /tag= b  
FT /note= "avermectin aglycon synthase protein"  
XX  
XX WO200050605-A1.  
XX  
XX 31-AUG-2000.  
XX  
XX 23-FEB-2000; 2000WO-JP001041.  
XX  
XX 24-FEB-1999; 99JP-00046961.  
XX  
XX (KITA ) KITASATO INST.  
XX  
XX Omura S, Ikeda H;  
XX  
XX WPI; 2000-565458/52.  
XX P-PSDB; AAB23751, AAB23752.  
XX  
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of  
XX it for the production of avermectin and its derivatives for drug and  
XX agrochemical use.  
XX  
XX Claim 2; Page 134-203; 314pp; Japanese.

XX The present sequence represents DNA which encodes avermectin aglycon  
XX synthase proteins. Also described are: (1) polypeptides encoded by all or  
XX part of the DNA; (2) expression vectors containing the DNA; (3) host

CC cells transformed by the vectors; (4) preparation of the polypeptides by  
CC culture of the transformants; (5) preparation of avermectin aglycon or  
CC its derivatives by culture of transformed avermectin-producing  
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length  
CC containing sense or antisense sequences from the avermectin aglycon  
CC synthase DNA. The enzymes are useful for the production of modified forms  
CC of avermectin and of the intermediates in its biosynthesis, for use as  
CC drugs, veterinary drugs and agrochemicals

XX SQ Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 U; 0 Other;

Query Match 4.6%; Score 23; DB 3; Length 31422;  
Best Local Similarity 100.0%; Pred. No. 0.64;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 AGCGTGATCTTCGACTACCCAC 215  
DB 4645 AGCGTGATCTTCGACTACCCAC 4667

## RESULT 10

AAH79278  
ID AAH79278 standard; DNA; 31422 BP.

XX AC AAH79278;

XX DT 04-DEC-2001 (first entry)

XX DE Streptomyces avermitilis coding sequences SEQ ID NO: 2.

XX KW Avermectin aglycone synthase; AAS; avermectin derivative;  
XX KW drug production; veterinary drug; pesticide; ds.

XX OS Streptomyces avermitilis.

XX FH Key Location/Qualifiers  
FT CDS 1. .14646  
FT /\*tag= a  
FT /product= "AAG65266"  
FT 14824. .31422  
FT /\*tag= b  
FT /product= "AAG65267"

XX WO200162939-A1.

XX PD 30-AUG-2001.

XX PF 23-FEB-2001; 2001WO-JP001381.

XX PR 24-FEB-2000; 2000JP-00047405.

XX XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX PA (KITA ) KITASATO INST.

XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;

XX WPI; 2001-582053/65.

XX DR P-FSD; AAG65266, AAG65267.

XX New modified avermectin aglycone synthase derived from Streptomyces  
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in  
PT drugs and pesticides.

XX PS Disclosure; Page 103-149; 257pp; Japanese.

XX The present invention relates to the production of modified derivatives  
CC of avermectin aglycone synthase (AAS) derived from Streptomyces  
CC avermitilis. The activity of an acyl carrier protein (ACP), beta-ketoacyl  
CC carrier protein synthase (KS), acyltransferase (AT), beta-ketoacyl  
CC carrier protein reductase (KR), dehydratase (DH), enoyl reductase (ER)  
CC and/or thioesterase (TE) domain may be reduced or suppressed. The process  
CC can be used in the production of drugs, veterinary drugs and pesticides.  
CC The present sequence is a fragment of the S. avermitilis genome

XX SQ Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 U; 0 Other;

Query Match 4.6%; Score 23; DB 4; Length 31422;

Best Local Similarity 100.0%; Pred. No. 0.64;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 AGCGTGATCTTCGACTACCCAC 215  
DB 4645 AGCGTGATCTTCGACTACCCAC 4667

## RESULT 11

AAI99682 18

Continuation (19 of 45) of AAI99682 from base 1800001 (Mycobacterium tuberculosis strai  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Query Match 4.6%; Score 23; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 0.55;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 AGCGTGATCTTCGACTACCCAC 215  
DB 81395 AGCGTGATCTTCGACTACCCAC 81417

## RESULT 12

AAI99683_18		Continuation (19 of 44) of AAI99683 from base 1800001 (Mycobacterium tuberculosis strain	
WP	Sequence split into 44 fragments	LOCUS AAI99683 Accession Aai99683	
WP	Fragment Name	Begin	End
WP	AAI99683_00	1	110000
WP	AAI99683_01	100001	210000
WP	AAI99683_02	200001	310000
WP	AAI99683_03	300001	410000
WP	AAI99683_04	400001	510000
WP	AAI99683_05	500001	610000
WP	AAI99683_06	600001	710000
WP	AAI99683_07	700001	810000
WP	AAI99683_08	800001	910000
WP	AAI99683_09	900001	1010000
WP	AAI99683_10	1000001	1110000
WP	AAI99683_11	1100001	1210000
WP	AAI99683_12	1200001	1310000
WP	AAI99683_13	1300001	1410000
WP	AAI99683_14	1400001	1510000
WP	AAI99683_15	1500001	1610000
WP	AAI99683_16	1600001	1710000
WP	AAI99683_17	1700001	1810000
WP	AAI99683_18	1800001	1910000
WP	AAI99683_19	1900001	2010000
WP	AAI99683_20	2000001	2110000
WP	AAI99683_21	2100001	2210000
WP	AAI99683_22	2200001	2310000
WP	AAI99683_23	2300001	2410000
WP	AAI99683_24	2400001	2510000
WP	AAI99683_25	2500001	2610000
WP	AAI99683_26	2600001	2710000
WP	AAI99683_27	2700001	2810000
WP	AAI99683_28	2800001	2910000
WP	AAI99683_29	2900001	3010000
WP	AAI99683_30	3000001	3110000
WP	AAI99683_31	3100001	3210000
WP	AAI99683_32	3200001	3310000
WP	AAI99683_33	3300001	3410000
WP	AAI99683_34	3400001	3510000
WP	AAI99683_35	3500001	3610000
WP	AAI99683_36	3600001	3710000
WP	AAI99683_37	3700001	3810000
WP	AAI99683_38	3800001	3910000
WP	AAI99683_39	3900001	4010000
WP	AAI99683_40	4000001	4110000
WP	AAI99683_41	4100001	4210000
WP	AAI99683_42	4200001	4310000
WP	AAI99683_43	4300001	4403765
Query Match 4.6%; Score 23; DB 4; Length 110000;			
Best Local Similarity 100.0%; Pred. No. 0.55;			
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	193	ACGCTGATCTTCGACTACCCAC	215
Db	72270	ACGCTGATCTTCGACTACCCAC	72292
RESULT 13			
ID	AAF30757		
XX	AAF30757	standard; DNA; 47981 BP.	
AC	AAF30757;		
XX	21-JUN-2001	(first entry)	
DE	Micromonospora megalomicea megalomicin biosynthetic gene cluster.		
XX	Megalomicin; meg gene; polyketide synthase; antibiotic; motilide;		
KW	antiparasitic; ds.		
XX	Micromonospora megalomicea subsp. nigra.		
OS			
XX			
FH	Key	Location/Qualifiers	
FT	CDS	complement(1..144)	
FT		/*tag= a	
FT		/partial	
FT		/gene= "megBVI(megT)" /product="TDP-4-keto-6-	
FT		deoxyglucose-"	
FT		/note= "encodes AAB82201"	
FT	CDS	928..2061	
FT		/*tag= b	
FT		/gene= "megDVI"	
FT		/product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"	
FT		/note= "encodes AAB82202"	
FT	CDS	2072..3382	
FT		/*tag= c	
FT		/gene= "megDI"	
FT		/product= "TDP-megosamine glycosyltransferase"	
FT		/note= "eryCIII homologue; encodes AAB802203"	
FT	CDS	3462..4634	
FT		/*tag= d	
FT		/gene= "megY"	
FT		/product= "mycarose O-acyltransferase"	
FT		/note= "encodes AAB82204"	
FT	CDS	4651..5775	
FT		/*tag= e	
FT		/gene= "megDII"	
FT		/product= "TDP-3-keto-6-deoxyhexose 3-aminotransaminase"	
FT		/note= "eryCI, DnrJ homologue, encodes AAB82205"	
FT	CDS	5822..6595	
FT		/*tag= f	
FT		/gene= "megDIII"	
FT		/product= "daunosaminyl-N,N-dimethyltransferase"	
FT		/note= "eryCVI homologue; encodes AAB82206"	
FT	CDS	6592..7197	
FT		/*tag= g	
FT		/gene= "megDIV"	
FT		/product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"	
FT		/note= "eryVIII, dnmU homologue, encodes AAB82207"	
FT	CDS	7220..8206	
FT		/*tag= h	
FT		/gene= "megDV"	
FT		/product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"	
FT		/note= "eryVIV, dnmV homologue; encodes AAB82208"	
FT	CDS	complement(8228..9220)	
FT		/*tag= i	
FT		/gene= "megBII-1(megDVII)"	
FT		/product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"	
FT		/note= "encodes AAB82209"	
FT	CDS	complement(9226..10479)	
FT		/*tag= j	
FT		/gene= "megBV"	
FT		/product= "TDP-mycarose glycosyltransferase"	
FT		/note= "encodes AAB82210"	
FT	CDS	complement(10483..11424)	
FT		/*tag= k	
FT		/gene= "megBIV"	
FT		/product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"	
FT		/note= "encodes AAB82211"	
FT	CDS	12181..22821	
FT		/*tag= l	
FT		/gene= "megA"	
FT		/product= "megalomicin 6-deoxyerithronolide B synthase 1"	
FT		/note= "polyketide synthase; encodes AAB82212"	
FT	misc_feature	12505..13470	
FT		/*tag= m	
FT		/gene= "megA"	
FT		/function= "AT-L"	
FT	misc_feature	13576..13791	
FT		/*tag= n	
FT		/gene= "megA"	
FT		/function= "ACP-L"	
FT	misc_feature	13849..15126	
FT		/*tag= o	
FT		/gene= "megA"	

```
FT FT /function= "KS1"
FT FT 15427. .16476
FT FT /*tag= p
FT FT /gene= "megA"
FT FT /function= "AT1"
FT FT 17155. .17694
FT FT /*tag= q
FT FT /gene= "megA"
FT FT /function= "KR1"
FT FT 17947. .18207
FT FT /*tag= r
FT FT /gene= "megA"
FT FT /function= "ACP1"
FT FT 18268. .19548
FT FT /*tag= s
FT FT /gene= "megA"
FT FT /function= "KS2"
FT FT 19876. .20910
FT FT /*tag= t
FT FT /gene= "megA"
FT FT /function= "AT2"
FT FT 21517. .22053
FT FT /*tag= u
FT FT /gene= "megA"
FT FT /function= "KR2"
FT FT 22318. .22575
FT FT /*tag= v
FT FT /gene= "megA"
FT FT /function= "ACP2"
FT FT 22867. .33555
FT FT /*tag= w
FT FT /gene= "megA"
FT FT /product= "megalomicin 6-deoxyerythronolide B synthase 2"
FT FT /note= "polyketide synthase, encodes AAB82213"
FT FT 22957. .24237
FT FT /*tag= x
FT FT /gene= "megAII"
FT FT /function= "KS3"
FT FT 22957. .24237
FT FT /*tag= ai
FT FT /gene= "megAII"
FT FT /function= "KS5"
FT FT 24544. .25581
FT FT /*tag= y
FT FT /gene= "megAII"
FT FT /function= "AT3"
FT FT 24544. .25581
FT FT /*tag= aj
FT FT /gene= "megAII"
FT FT /function= "AT5"
FT FT 26230. .26733
FT FT /*tag= z
FT FT /gene= "megAII"
FT FT /function= "KR3 (inactive)"
FT FT 26230. .26733
FT FT /*tag= ak
FT FT /gene= "megAII"
FT FT /function= "KR5"
FT FT 26998. .27258
FT FT /*tag= aa
FT FT /gene= "megAII"
FT FT /function= "ACP3"
FT FT 26998. .27258
FT FT /*tag= al
FT FT /gene= "megAII"
FT FT /function= "ACP5"
FT FT 27393. .28590
FT FT /*tag= ab
FT FT /gene= "megAII"
FT FT /function= "KS4"
FT FT 27393. .28590
FT FT /*tag= am
FT FT /gene= "megAII"
```

```
FT FT /function= "KS6"
FT FT 28897. .29931
FT FT /*tag= ac
FT FT /gene= "megAII"
FT FT /function= "AT4"
FT FT 28897. .29931
FT FT /*tag= an
FT FT /gene= "megAII"
FT FT /function= "AT6"
FT FT 29953. .30477
FT FT /*tag= ad
FT FT /gene= "megAII"
FT FT /function= "DH4"
FT FT 29953. .30477
FT FT /*tag= ao
FT FT /gene= "megAII"
FT FT /function= "KR6"
FT FT 31396. .32244
FT FT /*tag= ae
FT FT /gene= "megAII"
FT FT /function= "ER4"
FT FT 31396. .32244
FT FT /*tag= ap
FT FT /gene= "megAII"
FT FT /function= "ACP6"
FT FT 32257. .32799
FT FT /*tag= af
FT FT /gene= "megAII"
FT FT /function= "KR4"
FT FT 32257. .32799
FT FT /*tag= aq
FT FT /gene= "megAII"
FT FT /function= "TE"
FT FT 33052. .33312
FT FT /*tag= ag
FT FT /gene= "megAII"
FT FT /function= "ACP4"
FT FT 33666. .43271
FT FT /*tag= ah
FT FT /gene= "megAII"
FT FT /product= "megalomicin 6-deoxyerythronolide B synthase 3"

Query Match 4.2%; Score 21; DB 4; Length 47981;
Best Local Similarity 100.0%; Pred.No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 TTCGACTCGCTGGCGGGGTC 144
|||
DB 22447 TTCGACTCGCTGGCGGGGTC 22467

RESULT 14
ACA42620
ID ACA42620 standard; DNA; 4551 BP.
XX
AC ACA42620;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #24277.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Pseudomonas aeruginosa.
XX
FN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
```

PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlseen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
XX P-PSDB; ABU38750.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 30490; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
XX prokaryotic essential genes. Note: The sequence data for this patent did  
XX not form part of the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 4551 BP; 779 A; 1608 C; 1519 G; 645 T; 0 U; 0 Other;  
XX  
XX Query Match 4.0%; Score 20; DB 7; Length 4551;  
XX Best Local Similarity 100.0%; Pred. No. 19;  
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 258 GGAGCCCGACGACGCGCTGG 277  
XX |||||  
XX Db 3600 GGAGCCCGACGACGCGCTGG 3619  
XX  
XX RESULT 15  
XX AAQ46806  
XX ID AAQ46806 standard; DNA; 29879 BP.  
XX  
XX AC AAQ46806;  
XX  
XX 27-AUG-2003 (revised)  
XX 25-MAR-2003 (revised)  
XX 22-DEC-1993 (first entry)  
XX  
XX eryA region of *S. erythraea* chromosome.

XX Saccharopolyspora erythraea; eryA; biosynthesis; polyketide; module;  
XX erythromycin; condensation; elongation; acyl chain growth;  
XX gene replacement; ss.  
XX  
XX Saccharopolyspora erythraea.  
XX  
XX OS  
XX FH Key Location/Qualifiers  
XX FT CDS 3..10199  
XX FT /\*tag= a  
XX FT /note= "ORF 1"  
XX FT 10218..20921  
XX FT /\*tag= b  
XX FT /note= "ORF 2"  
XX FT 20922..29879  
XX FT /\*tag= C  
XX FT /note= "ORF 3"  
XX  
XX PN WO9313663-A1.  
XX  
XX PD 22-JUL-1993.  
XX  
XX PP 17-JAN-1992; 92WO-US0000427.  
XX  
XX PR 17-JAN-1992; 92WO-US0000427.  
XX  
XX PA (ABBO ) ABBOTT LAB.  
XX  
XX PI Katz L, Donadio S, Mcalpine JB;  
XX WPI; 1993-242804/30.  
XX P-PSDB; AAR44430, AAR44431, AAR44432.  
XX  
XX Biosynthesis of specific polyketide analogues esp. erythromycin cpds. -  
XX by introducing altered biosynthetic gene-contg. DNA into microorganisms.  
XX  
XX Claim 27; Fig 2; 133pp; English.  
XX  
XX This sequence represents a fragment of the Saccharopolyspora erythraea  
XX genome, designated eryA. The polypeptides encoded by this region are  
XX involved in the biosynthesis of the polyketide segment of erythromycin.  
XX eryA is organised in modules and each module takes care of one  
XX condensation step. The precise succession of elongation steps is dictated  
XX by the genetic order of the modules. This fragment may be specifically  
XX altered such that novel polyketide molecules of desired structure are  
XX produced. Three types of alteration may be produced; those inactivating a  
XX single function in a module which does not arrest acyl chain growth;  
XX those inactivating a single function in a module which does affect chain  
XX growth; and those affecting an entire module. The mutations may be  
XX introduced by gene replacement. (Updated on 25-MAR-2003 to correct PN  
XX field.) (Updated on 27-AUG-2003 to correct OS field.)  
XX  
XX Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 U; 0 Other;  
XX  
XX Query Match 4.0%; Score 20; DB 2; Length 29879;  
XX Best Local Similarity 100.0%; Pred. No. 15;  
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 124 TTCGACTCGTGGCGCGGT 143  
XX |||||  
XX Db 20553 TTCGACTCGTGGCGCGGT 20572  
XX  
XX RESULT 16  
XX AAQ92302/c  
XX ID AAQ92302 standard; DNA; 31422 BP.  
XX  
XX AC AAQ92302;  
XX  
XX 10-JAN-2001 (first entry)  
XX  
XX S. avermitilis avermectin aglycon synthase DNA aveAI1 SEQ ID NO:2.  
XX

KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;  
 KW multifunctional enzyme; polyketide; avermectin; veterinary drug;  
 KW agrochemical; ds.  
 XX OS Streptomyces avermitilis.  
 XX Key Location/Qualifiers  
 FH CDS 1. .14646  
 FT /tag= a  
 FT /product= "AAG65266"  
 FT 14824. .31422  
 FT /tag= b  
 FT /product= "AAG65267"  
 XX WO200050605-A1.  
 XX 31-AUG-2000.  
 XX 23-FEB-2000; 2000WO-JP001041.  
 XX 24-FEB-1999; 99JP-00046961.  
 XX (KITA ) KITASATO INST.  
 XX Omura S, Ikeda H;  
 XX WPI; 2000-565458/52.  
 XX P-PSDB; AAB23751, AAB23752.  
 XX Avermectin aglycone synthase DNA and proteins encoded by all or part of  
 PT it for the production of avermectin and its derivatives for drug and  
 PT agrochemical use.  
 XX Claim 2; Page 134-203; 314pp; Japanese.  
 XX The present sequence represents DNA which encodes avermectin aglycon  
 CC synthase proteins. Also described are: (1) polypeptides encoded by all or  
 CC part of the DNA; (2) expression vectors containing the DNA; (3) host  
 CC cells transformed by the vectors; (4) preparation of the polypeptides by  
 CC culture of the transformants; (5) preparation of avermectin aglycon or  
 CC its derivatives by culture of transformed avermectin-producing  
 CC microorganisms; and (6) oligonucleotides of 5-60 bases in length  
 CC containing sense or antisense sequences from the avermectin aglycon  
 CC synthase DNA. The enzymes are useful for the production of modified forms  
 CC of avermectin and of the intermediates in its biosynthesis, for use as  
 CC drugs, veterinary drugs and agrochemicals  
 XX SQ Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 U; 0 Other;  
 Query Match 4.0%; Score 20; DB 3; Length 31422;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 366 CCTCGCCGACACCGGCACCG 385  
 Db 27210 CCTCGCCGACACCGGCACCG 27191  
 RESULT 17  
 AAH79278/c  
 ID AAH79278 standard; DNA; 31422 BP.  
 XX AC AAH79278;  
 XX 04-DEC-2001 (first entry)  
 XX Streptomyces avermitilis coding sequences SEQ ID NO: 2.  
 DE DE  
 XX Avermectin aglycone synthase; AAS; avermectin derivative;  
 KW drug production; veterinary drug; pesticide; ds.  
 XX OS Streptomyces avermitilis.  
 XX

FH Key Location/Qualifiers  
 FT CDS 1. .14646  
 FT /tag= a  
 FT /product= "AAG65266"  
 FT 14824. .31422  
 FT /tag= b  
 FT /product= "AAG65267"  
 XX WO200162939-A1.  
 XX 30-AUG-2001.  
 XX 23-FEB-2001; 2001WO-JP001381.  
 XX 24-FEB-2000; 2000JP-00047405.  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX (KITA ) KITASATO INST.  
 XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;  
 XX WPI; 2001-582053/65.  
 XX P-PSDB; AAG65266, AAG65267.  
 XX New modified avermectin aglycone synthase derived from Streptomyces  
 PT avermectilis used in production of 22,23-dihydroavermectin B1a used in  
 PT drugs and pesticides.  
 XX Disclosure; Page 103-149; 257pp; Japanese.  
 XX The present invention relates to the production of modified derivatives  
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces  
 CC avermitilis. The activity of an acyl carrier protein (ACP), beta-ketoacyl  
 CC carrier protein synthase (KS), acyltransferase (AT), beta-ketoacyl  
 CC carrier protein reductase (KR), dehydratase (DH), enoyl reductase (ER)  
 CC and/or thioesterase (TE) domain may be reduced or suppressed. The process  
 CC can be used in the production of drugs, veterinary drugs and pesticides.  
 CC The present sequence is a fragment of the S. avermitilis genome  
 XX SQ Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 U; 0 Other;  
 Query Match 4.0%; Score 20; DB 4; Length 31422;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 366 CCTCGCCGACACCGGCACCG 385  
 Db 27210 CCTCGCCGACACCGGCACCG 27191  
 RESULT 18  
 AAT80413  
 ID AAT80413 standard; DNA; 43280 BP.  
 XX AC AAT80413;  
 XX 27-FEB-1998 (first entry)  
 XX Tyllactone synthase gene cluster.  
 XX Tyllactone synthase gene cluster;  
 KW Tyllactone synthase gene cluster; tyIG gene; multifunctional protein;  
 KW polyketide; tyllactone synthesis; antibiotic; tylosin; ss.  
 XX OS Streptomyces fradiae.  
 XX Key Location/Qualifiers  
 FH CDS 816. .14243  
 FT /tag= a  
 FT /transl\_except= (pos: 816. .818, aa: Met)  
 FT /note= "ORF1 encodes protein shown in AAW22601"  
 FT 14351. .19945  
 FT /tag= b  
 FT /transl\_except= (pos: 14351. .14353, aa: Met)  
 FT





PT New peptide useful as a marker for the diagnosis of breast cancer.

PS Claim 1; Page 272; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides  
 CC (AA07544-AAL26789) and methods of assessing whether a patient is  
 CC afflicted with breast cancer by examining the correlation between the  
 CC expression of certain markers and the cancerous state of breast cells.  
 CC The polynucleotides and encoded polypeptides are potential markers for  
 CC detecting, diagnosing, monitoring, characterising treating and  
 CC potentially preventing breast cancer. The polynucleotides and encoded  
 CC polypeptides are also useful for isolating compounds with cytostatic  
 CC activity

SQ Sequence 177 BP; 35 A; 56 C; 49 G; 37 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 4; Length 177;  
 Best Local Similarity 100.0%; Pred. No. 79;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCGAGCAGACCGCATC 30  
 |||||  
 Db 157 CGCGAGCAGACCGCATC 139

RESULT 21

AAAT73432  
 ID AAAT73432 standard; DNA; 375 BP.

XX AC AAAT73432;

XX 07-OCT-1997 (first entry)

DE Human precursor urocortin coding sequence.

XX Human; precursor; urocortin; corticotropin releasing factor; receptor;  
 KW adrenocorticotrophic hormone; ACTH; affinity; rat; ovine; carp; urotensin;  
 KW placenta; probe; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FT mat\_peptide 247..366  
 FT /\*tag= a

XX WO9700063-A2.

XX 03-JAN-1997.

XX 12-JUN-1996; 96WO-US010240.

XX 13-JUN-1995; 95US-0028444P.

XX 13-JUN-1995; 95US-00490314.

XX 11-AUG-1995; 95US-0002223P.

XX (SALK ) SALK INST BIOLOGICAL STUDIES.

XX Vale WW, Vaughan J, Donaldson CJ, Lewis KA, Sawchenko P;

XX Rivier JEF, Perrin MH;

XX WPI; 1997-077344/07.

XX P-PSDB; AAW22392.

XX Urocortin peptide(s) related to urotensin and corticotropin-releasing  
 PT factor - for increasing ACTH and beta-endorphin levels, lowering blood  
 PT pressure and improving mood, memory and learning performance.

XX Claim 13; Page 82; 91pp; English.

XX This nucleotide sequence encodes a human precursor urocortin peptide  
 CC which: (i) binds to corticotropin-releasing factor (CRF) receptor 2 (CRF-  
 CC R2); (ii) promotes the production of adrenocorticotrophic hormone (ACTH);  
 CC (iii) exhibits a higher affinity for the long form of CRF-R2 than does

CC rat/human CRF (r/hCRF); and (iv) has < 80% homology with r/hCRF, ovine  
 CC CRF or carp urotensin. The sequence was isolated from a human genomic  
 CC placental DNA library using a 160 bp fragment of the corresponding rat  
 CC gene (AAAT73433) as a probe. The nucleotide sequence encoding the mature  
 CC human urocortin peptide has 88% similarity to the rat sequence (AAAT73433)

XX Sequence 375 BP; 52 A; 128 C; 145 G; 50 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 72;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCGAGCAGACCGCATC 30  
 |||||

Db 333 CGCGAGCAGACCGCATC 351

RESULT 22

ACA63335  
 ID ACA63335 standard; DNA; 375 BP.

XX AC ACA63335;

XX 29-AUG-2003 (first entry)

XX DNA encoding human urocortin.

XX Human; urocortin; adrenocorticotrophic hormone; ACTH; blood flow; memory;  
 KW central nervous system disorder; blood pressure; vascular bed; anxiety;  
 KW coronary blood flow; inflammation; vascular permeability; parturition;  
 KW CRF-binding protein; hypothalamic pituitary adrenal function; swelling;  
 KW Alzheimer's disease; chronic fatigue syndrome; appetite; alertness; ds;  
 KW respiratory system; learning performance; depression; endocrine disorder;  
 KW gene.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FT CDS 1..375  
 FT /\*tag= a  
 FT /product= "Urocortin"

XX US2003032587-A1.

XX 13-FEB-2003.

XX 26-MAR-2001; 2001US-00818009.

XX 13-JUN-1995; 95US-0028444P.

XX 11-AUG-1995; 95US-0002223P.

XX 12-JUN-1996; 96WO-US010240.

XX 10-DEC-1997; 97US-00981189.

XX (SALK ) SALK INST BIOLOGICAL STUDIES.

XX Valle WW, Vaughan J, Donaldson CJ, Lewis KA, Sawchenko P;

XX Rivier JEF, Perrin MH;

XX WPI; 1997-077344/07.

XX P-PSDB; ABU62366.

XX Urocortin peptide(s) related to urotensin and corticotropin-releasing  
 PT factor - for increasing ACTH and beta-endorphin levels, lowering blood  
 PT pressure and improving mood, memory and learning performance.

XX Disclosure; Page 7; 34pp; English.

XX The invention relates to a human urocortin (Ucn) peptide or an analogous  
 CC sequence having only conservative substitutions to the amino acid  
 CC residues in it, or an N-terminally shortened fragment of either which is  
 CC biologically active to increase adrenocorticotrophic hormone (ACTH)  
 CC production. Human urocortin or its N-terminally shortened antagonist  
 CC peptide are useful for modifying blood flow and/or blood pressure and is

CC further useful for modulating blood flow in a desired vascular bed. Human  
 CC urocortin is also useful for increasing coronary blood flow and for  
 CC decreasing swelling and/or inflammation and/or vascular permeability. A  
 CC CRF-binding protein blocking compound is useful for increasing the in  
 CC vivo level of CRF and/or Ucn. The amount of CRF-binding protein blocking  
 CC compound is sufficient to promote parturition in a pregnant female. The  
 CC amount of the compound administered is effective so as to result in an  
 CC increase in free endogenous CRF and/or Ucn in the brain which causes  
 CC improvement in short to medium term memory in a subject afflicted with  
 CC Alzheimer's disease, relief from chronic fatigue syndrome, suppression of  
 CC appetite, stimulation of the respiratory system, improvement in learning  
 CC performance, improvement in memory, improvement in alertness, reduction  
 CC of depression and/or lessening of anxiety. The compound is administered  
 CC so that it reaches the brain. Human urocortin is useful for evaluating  
 CC hypothalamic pituitary adrenal function in mammals with suspected  
 CC endocrine or central nervous system pathology. Human urocortin antibodies  
 CC are useful in diagnostic methods and systems for detecting the level of  
 CC ucn polypeptide, for immunoaffinity or affinity chromatography  
 CC purification of Ucn, and also in human therapeutic methods. The present  
 CC sequence represents DNA encoding human urocortin

SQ Sequence 375 BP; 52 A; 128 C; 145 G; 50 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGACCGCATC 30  
 |||||  
 Db 333 CGCCGAGCAGACCGCATC 351

## RESULT 23

AAZ35707  
 ID AAZ35707 standard; cDNA; 375 BP.

AC AAZ35707;

DT 31-JAN-2000 (first entry)

DE Human urocortin encoding cDNA.

XX Human; urocortin; protection; ischaemic injury; cardiac myocyte;  
 KW cardioprotective; Cardiac ischaemia; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FT CDS 1..375  
 FT /tag= a  
 FT /product= "urocortin"

FN WO951261-A1.

PD 14-OCT-1999.

XX 07-APR-1999; 99WO-GH001056.

XX 07-APR-1998; 98GB-00007504.

XX (UNLO ) UNIV COLLEGE LONDON.

XX Stephanou A, Brar B, Knight R, Latchman D;

XX WPI; 2000-013030/01.

XX P-PSDB; AAY49943.

XX Use of urocortin or cardio-protective derivative, useful for preparation  
 PT of medicament for treating cardiac ischaemia.

XX Disclosure; Page 35-36; 42pp; English.

XX The present sequence encodes human urocortin. The present invention

CC describes the use of urocortin or its cardio-protective derivative, in  
 CC the manufacture of a medicament for use in the treatment of cardiac  
 CC ischaemia. Also described are: (1) a method for the treatment of a  
 CC patient who has suffered a cardiac ischaemia, comprising administering  
 CC urocortin to the patient; (2) urocortin for use in a method of treatment  
 CC of the human or animal body by therapy; and (3) an agent for treating  
 CC cardiac ischaemia comprising urocortin

SQ Sequence 375 BP; 52 A; 128 C; 145 G; 50 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 3; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGACCGCATC 30  
 |||||  
 Db 333 CGCCGAGCAGACCGCATC 351

## RESULT 24

AAV99120  
 ID AAV99120 standard; DNA; 452 BP.

XX AAV99120;

DT 09-MAR-1999 (first entry)

DE DNA methyltransferase genomic fragment.

XX DNA methyltransferase; DNA Mefase; antisense oligonucleotide; human;  
 KW cellular growth; tumour growth inhibition; silenced gene activation;  
 KW beta thalassemia; sickle cell anemia; ss.

XX Homo sapiens.

XX WO9854313-A2.

XX 03-DEC-1998.

XX 29-MAY-1998; 98WO-IB001107.

XX 30-MAY-1997; 97US-00866340.

XX 17-DEC-1997; 97US-0069865P.

XX (UYMC-) UNIV MCGILL.

XX Szyf M, Bigey P, Ramchandani S;

XX WPI; 1999-059833/05.

XX New DNA methyltransferase nucleotide sequences - used particularly to  
 PT develop antisense oligonucleotides for diagnostic and therapeutic  
 PT purposes, particularly for inhibiting tumour growth.

XX Claim 1; Page 59; 108pp; English.

XX AAV99091-128 represent DNA methyltransferase (DNA Mefase) genomic  
 CC fragments. The sequences together sequentially produce the full length  
 CC sequence AAV99128. The sequences can be used to derive antisense  
 CC oligonucleotides which inhibit DNA Mefase expression. The antisense  
 CC oligonucleotides can be used in investigating the role of DNA Mefase in  
 CC cellular growth. They can be administered at different points in the cell  
 CC cycle, or in conjugation with promoters or inhibitors of cell growth to  
 CC determine the role of DNA Mefase in the growth of the cell type of  
 CC interest. The antisense oligonucleotides can also be used for inhibiting  
 CC tumour growth in a mammal, or to activate silenced genes to provide a  
 CC missing gene function. This ameliorates disease symptoms, e.g. in beta  
 CC thalassemia and sickle cell anemia. The antisense oligonucleotides can  
 CC also be used as analytical and diagnostic tools and a potentiators of  
 CC transgenic plant and animal studies

XX Sequence 452 BP; 100 A; 135 C; 131 G; 86 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 452;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 CATCTGCTGAAGCTGGTC 45  
|||||  
DB 201 CATCTGCTGAAGCTGGTC 219  
|||||

RESULT 25  
ABS63429  
ID ABS63429 standard; cDNA; 953 BP.  
XX  
AC ABS63429;  
XX  
DT 15-NOV-2002 (first entry)  
XX  
DE DNA encoding wheat caffeoyl-CoA O-methyltransferase (CCOMT).  
XX  
KW Plant; caffeoyl-CoA O-methyltransferase; CCOMT; transgenic; corn; rice;  
KW lignin; soybean; wheat; gene; ss.  
XX  
OS Triticum aestivum.  
XX  
PN US2002081693-A1.  
XX  
PD 27-JUN-2002.  
XX  
PF 01-DEC-1999; 99US-00452239.  
XX  
PR 02-DEC-1998; 98US-0110594P.  
XX  
PA (CAHO/) CAHOON R E.  
PA (FADE/) FADER G M.  
PA (RAFA/) RAFALSKI J A.  
XX  
PI Cahoon RB, Fader GM, Rafalski JA;  
XX  
DR WPI; 2002-626543/67.  
DR P-PSDB; ABG78357.

Novel plant caffeoyl-CoA O-methyltransferase polypeptide and polynucleotide encoding it useful for producing transgenic plants having altered level of lignin production in cells.

Claim 2; Page 48-49; 59pp; English.

The invention relates to novel isolated plant caffeoyl-CoA O-methyltransferase (CCOMT) polypeptides (I) and encoding polynucleotides (II). (I) is useful for evaluating at least one compound for its ability to inhibit the activity of a caffeoyl-CoA O-methyltransferase. (II) is useful as a probe or primer for obtaining a nucleic acid fragment encoding CCOMT polypeptide by hybridisation and amplification. (II) or a polynucleotide comprising at least 30 contiguous nucleotides derived from (II), is useful for selecting an isolated polynucleotide that affects the level of expression of methyltransferase polypeptide in a plant cell. (II) is useful for positive selection of a transformed cell, by transforming a plant cell, preferably a monocot or dicot, with (II) and growing the transformed plant cell under conditions allowing expression of the polynucleotide to complement CCOMT auxotroph to provide a positive selection. (I) is useful for preparing antibodies for use in detecting CCOMT polypeptides in situ in cells or in vitro in cell extracts. (II) or its fragments are useful as probes for genetically and physically mapping the genes that they are part of and as markers for traits linked to those genes. Fragments of (II) are useful to create transgenic plants in which CCOMT polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This is useful for altering the level of lignin production in the plant cells. (II) is also useful in direct fluorescence in situ hybridisation mapping. ABS63407-ABS63430 represent plant CCOMT coding sequences of the invention

Sequence 953 BP; 192 A; 336 C; 279 G; 146 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 6; Length 953;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 CTGCTGAAGCTGGTCGGCG 49  
|||||  
DB 591 CTGCTGAAGCTGGTCGGCG 609  
|||||

RESULT 26  
ABS63428  
ID ABS63428 standard; cDNA; 1049 BP.  
XX  
AC ABS63428;  
XX  
DT 15-NOV-2002 (first entry)  
XX  
DE DNA encoding wheat caffeoyl-CoA O-methyltransferase (CCOMT).  
XX  
KW Plant; caffeoyl-CoA O-methyltransferase; CCOMT; transgenic; corn; rice;  
KW lignin; soybean; wheat; gene; ss.  
XX  
OS Triticum aestivum.  
XX  
PN US2002081693-A1.  
XX  
PD 27-JUN-2002.  
XX  
PF 01-DEC-1999; 99US-00452239.  
XX  
PR 02-DEC-1998; 98US-0110594P.  
XX  
PA (CAHO/) CAHOON R E.  
PA (FADE/) FADER G M.  
PA (RAFA/) RAFALSKI J A.  
XX  
PI Cahoon RB, Fader GM, Rafalski JA;  
XX  
DR WPI; 2002-626543/67.  
DR P-PSDB; ABG78356.

Novel plant caffeoyl-CoA O-methyltransferase polypeptide and polynucleotide encoding it useful for producing transgenic plants having altered level of lignin production in cells.

Claim 2; Page 47; 59pp; English.

The invention relates to novel isolated plant caffeoyl-CoA O-methyltransferase (CCOMT) polypeptides (I) and encoding polynucleotides (II). (I) is useful for evaluating at least one compound for its ability to inhibit the activity of a caffeoyl-CoA O-methyltransferase. (II) is useful as a probe or primer for obtaining a nucleic acid fragment encoding CCOMT polypeptide by hybridisation and amplification. (II) or a polynucleotide comprising at least 30 contiguous nucleotides derived from (II), is useful for selecting an isolated polynucleotide that affects the level of expression of methyltransferase polypeptide in a plant cell. (II) is useful for positive selection of a transformed cell, by transforming a plant cell, preferably a monocot or dicot, with (II) and growing the transformed plant cell under conditions allowing expression of the polynucleotide to complement CCOMT auxotroph to provide a positive selection. (I) is useful for preparing antibodies for use in detecting CCOMT polypeptides in situ in cells or in vitro in cell extracts. (II) or its fragments are useful as probes for genetically and physically mapping the genes that they are part of and as markers for traits linked to those genes. Fragments of (II) are useful to create transgenic plants in which CCOMT polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This is useful for altering the level of lignin production in the plant cells. (II) is also useful in direct fluorescence in situ hybridisation mapping. ABS63407-ABS63430 represent plant CCOMT coding sequences of the invention

XX	cDNA encoding an acetylcholinesterase protein.
DE	Acetylcholinesterase; AchE; fusion protein; ligand receptor; monomer;
XX	ligand detection; marker enzyme; ss.
KW	Bungarus fasciatus.
OS	
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..1698
FT	/tag= a
FT	/transl_except= (pos: 52..54, aa: Leu)
FT	/transl_except= (pos: 55..57, aa: Cys)
FT	/transl_except= (pos: 58..60, aa: Ile)
FT	/transl_except= (pos: 61..63, aa: Pro)
FT	/transl_except= (pos: 64..66, aa: Ser)
FT	/transl_except= (pos: 67..69, aa: Cys)
FT	/transl_except= (pos: 70..72, aa: Val)
FT	/transl_except= (pos: 73..75, aa: Ala)
FT	/transl_except= (pos: 76..78, aa: Val)
FT	/transl_except= (pos: 79..81, aa: Leu)
FT	/transl_except= (pos: 82..84, aa: Gln)
FT	/transl_except= (pos: 85..87, aa: Leu)
FT	/product= "acetylcholinesterase"
FT	/note= "no termination codon given"
XX	
PN	FR2773802-A1.
XX	
PD	23-JUL-1999.
XX	
PF	22-JAN-1998; 98FR-00000656.
XX	
PR	22-JAN-1998; 98FR-00000656.
XX	(INSP ) INST PASTEUR.
PA	(INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX	
PI	Bon C, Choumet V, Cousin X;
XX	
DR	WPI; 1999-471239/40.
DR	P-PSDB; AAY30100.
XX	
XX	A fusion protein comprising an acetyl cholinesterase and ligand receptor,
FT	useful for detection of ligands.
XX	
PS	Claim 19; Page 98-99; 114pp; French.
XX	
CC	The present sequence encodes an acetylcholinesterase protein (AChE). The
CC	protein is used to make a fusion protein comprising an AChE monomer and a
CC	specific ligand receptor. The AChE fusion protein is useful for the
CC	production of an AChE monomer in a soluble format. The AChE fusion
CC	polypeptide is useful for detection of ligands in samples. AChE is used
CC	as a marker enzyme, in a similar manner to peroxidase, alkaline
CC	phosphatase and beta-galactosidase. By having AChE fused to a receptor
CC	protein, various ligands can be detected by their binding to the receptor
CC	portion of the fusion polypeptide
XX	
SQ	Sequence 1698 BP; 305 A; 595 C; 495 G; 303 T; 0 U; 0 Other;
	Query Match 3.8%; Score 19; DB 2; Length 1698;
	Best Local Similarity 100.0%; Pred. No. 60;
	Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	407 CCACCGAAGCGGCCGGC 425
DB	712 CCACCGAAGCGGCCGGC 694
	RESULT 29
	AAS54382
ID	AAS54382 standard; DNA; 1893 BP.
XX	
AC	AAS54382;

XX DT 13-FEB-2002 (first entry)  
 XX DE Pseudomonas aeruginosa DNA for cellular proliferation protein #513.  
 XX KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;  
 XX KW antibacterial; drug design.  
 XX OS Pseudomonas aeruginosa.  
 XX PN WO200170955-A2.  
 XX PD 27-SEP-2001.  
 XX PF 21-MAR-2001; 2001WO-US003180.  
 XX PR 21-MAR-2000; 2000US-0191078P.  
 XX PR 23-MAY-2000; 2000US-0206848P.  
 XX PR 26-MAY-2000; 2000US-0207727P.  
 XX PR 23-OCT-2000; 2000US-0242578P.  
 XX PR 27-NOV-2000; 2000US-0253625P.  
 XX PR 22-DEC-2000; 2000US-0257931P.  
 XX PR 16-FEB-2001; 2001US-0269308P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 XX PI Yamamoto RT, Xu HH;  
 XX WPI; 2001-611495/70.  
 XX P-PSDB; AAU36523.  
 XX PT New polynucleotides for the identification and development of  
 XX PT antibiotics, comprise sequences of antisense nucleic acids.  
 XX PS Claim 27; SEQ ID NO 8019; 511pp; English.  
 XX CC The invention relates to antisense inhibitors of genes essential to  
 XX CC prokaryotic cellular proliferation, their use in identifying the genes,  
 XX CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
 XX CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
 XX CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
 XX CC useful for the identification of potential new targets for antibiotic  
 XX CC development. The antisense nucleic acids can also be used to identify  
 XX CC proteins used in proliferation, to express these proteins, and to obtain  
 XX CC antibodies capable of binding to the expressed proteins. The proteins can  
 XX CC be used to screen compounds in rational drug discovery programmes. The  
 XX CC antisense nucleic acid sequence is also useful to screen for homologous  
 XX CC nucleic acids which are required for cell proliferation in a wide variety  
 XX CC of organisms. The present sequence encodes an essential prokaryotic  
 XX CC cellular proliferation protein. Note: The sequence data for this patent  
 XX CC did not form part of the printed specification, but was obtained in  
 XX CC electronic format directly from WIPO at  
 XX CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 1893 BP; 374 A; 627 C; 575 G; 317 T; 0 U; 0 Other;  
 Query Match 3.8%; Score 19; DB 4; Length 1893;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 84 CGAAGGCATCGCCGCGC 102  
 DB 807 CGAAGGCATCGCCGCGC 925  
 RESULT 30  
 ACA42802  
 ID ACA42802 standard; DNA; 1893 BP.  
 XX AC ACA42802;  
 XX

DT 19-JUN-2003 (first entry)  
 XX DE Prokaryotic essential gene #24459.  
 XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 XX KW drug design; gene.  
 XX OS Pseudomonas aeruginosa.  
 XX PN WO200277183-A2.  
 XX PD 03-OCT-2002.  
 XX PF 21-MAR-2002; 2002WO-US009107.  
 XX PR 21-MAR-2001; 2001US-00815242.  
 XX PR 06-SEP-2001; 2001US-00948993.  
 XX PR 25-OCT-2001; 2001US-0342923P.  
 XX PR 08-FEB-2002; 2002US-00072851.  
 XX PR 06-MAR-2002; 2002US-0362699P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 XX P-PSDB; ABU38932.  
 XX PT New antisense nucleic acids, useful for identifying proteins or screening  
 XX PT for homologous nucleic acids required for cellular proliferation to  
 XX PT isolate candidate molecules for rational drug discovery programs.  
 XX PS Claim 14; SEQ ID NO 30672; 1766pp; English.  
 XX CC The invention relates to an isolated nucleic acid comprising any one of  
 XX CC the 6213 antisense sequences given in the specification where expression  
 XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 XX CC encoding a polypeptide whose expression is inhibited by the antisense  
 XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 XX CC polypeptide or its fragment whose expression is inhibited by the  
 XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 XX CC proliferation or the activity of a gene in an operon required for  
 XX CC proliferation; (7) identifying a compound that influences the activity of  
 XX CC the gene product or that has an activity against a biological pathway  
 XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
 XX CC identifying a gene required for cellular proliferation or the biological  
 XX CC pathway in which a proliferation-required gene or its gene product lies  
 XX CC or a gene on which the test compound that inhibits proliferation of an  
 XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 XX CC compound's activity; (11) a culture comprising strains in which the gene  
 XX CC product is overexpressed or underexpressed; (12) determining the extent  
 XX CC to which each of the strains is present in a culture or collection of  
 XX CC strains; or (13) identifying the target of a compound that inhibits the  
 XX CC proliferation of an organism. The antisense nucleic acids are useful for  
 XX CC identifying proteins or screening for homologous nucleic acids required  
 XX CC for cellular proliferation to isolate candidate molecules for rational  
 XX CC drug discovery programs, or for screening homologous nucleic acids  
 XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 XX CC prokaryotic essential genes. Note: The sequence data for this patent did  
 XX CC not form part of the printed specification, but was obtained in  
 XX CC electronic format directly from WIPO at  
 XX CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 1893 BP; 374 A; 627 C; 575 G; 317 T; 0 U; 0 Other;  
 Query Match 3.8%; Score 19; DB 7; Length 1893;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CGAAGGCATCGCCCGGC 102  
DB 807 CGAAGGCATCGCCCGGC 825  
RESULT 31  
ABN85313/c  
ID ABN85313 standard; cDNA; 2040 BP.  
XX AC ABN85313;  
XX DT 30-SEP-2002 (first entry)  
XX DE Human cytoskeleton-associated protein, CSAP-4, coding sequence.  
XX KW Human; cytoskeleton-associated protein; CSAP; CSAP-4;  
KW cell proliferative disorder; viral infection; neurological disorder;  
KW transgenic animal; antiatherosclerotic; antiproliferative; antiinflammatory;  
KW virucide; anticonvulsant; vasotropic; cerebroprotective; neurotropic;  
KW neuroprotective; cytostatic; gene therapy; gene; ss.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX CDS 291..1271  
XX FT /\*tag= a  
XX ET /product= "CSAP-4"  
XX FN WO200253719-A2.  
XX PD 11-JUL-2002.  
XX PF 04-JAN-2002; 2002WO-US000178.  
XX PR 04-JAN-2001; 2001US-0260085P.  
XX PR 13-FEB-2001; 2001US-0268554P.  
XX PR 14-FEB-2001; 2001US-0269111P.  
XX PR 23-FEB-2001; 2001US-0271211P.  
XX KW (INCY-) INCYTE GENOMICS INC.  
XX LU DAM, Baughn MR, Yao MG, Ding L, Honchell CD, Yue H, Tang YT;  
PI Warren BA, Duggan BM, Xu Y, Walla NK, Griffin JA, Stewart BA;  
PI Gandhi AR, Khan PA, Thangavelu K, Ison CH, Azimzai Y, Hafalia AJA;  
PI Gieczen KJ, Lal PG, Sanjanwala NM, Elliott VS;  
XX WPI; 2002-583611/62.  
XX DR P-PSDB; ABB83475.  
XX PT Novel isolated human cytoskeleton-associated protein for diagnosing,  
PT treating or preventing atherosclerosis, psoriasis, leukemia, epilepsy,  
PT ischemic cerebrovascular disease, cerebral neoplasms and Alzheimer's  
PT disease.  
XX PS Claim 5; Page 154-155; 167pp; English.  
XX CC The present sequence is the coding sequence for a human cytoskeleton-  
CC associated protein (CSAP). The CSAP and its coding sequence are useful in  
CC the diagnosis, treatment and prevention of a cell proliferative disorder  
CC such as actinic keratosis, atherosclerosis, psoriasis, primary  
CC thrombocythaemia, leukaemia; a viral infection such as those caused by  
CC adenoviruses (acute respiratory disease, pneumonia), arenaviruses  
CC (lymphocytic choriomeningitis); and a neurological disorder such as  
CC epilepsy, ischaemic cerebrovascular disease, stroke, cerebral neoplasms,  
CC Alzheimer's disease, Pick's disease, Huntington's disease or amyotrophic  
CC lateral sclerosis. The CSAP coding sequence is also useful for creating  
CC knock out or knock in humanised animals or transgenic animals to model  
CC human diseases  
XX SQ Sequence 2040 BP; 465 A; 584 C; 650 G; 341 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 6; Length 2040;  
Best Local Similarity 100.0%; Pred. No. 59;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 12 CGCCGAGCAGAACCGCATC 30  
DB 1949 CGCCGAGCAGAACCGCATC 1931  
RESULT 32  
ADD29815  
ID ADD29815 standard; mRNA; 2448 BP.  
XX AC ADD29815;  
XX DT 15-JAN-2004 (first entry)  
XX DE Human tumour suppressor mRNA SEQ ID NO:311.  
XX KW ss; human; tumour suppressor; cancer; cancer; cytostatic; gene therapy.  
XX OS Homo sapiens.  
XX FN WO2003058201-A2.  
XX PD 17-JUL-2003.  
XX PF 31-DEC-2002; 2002WO-US041825.  
XX PR 31-DEC-2001; 2001US-0345317P.  
XX KW (QUAR-) QUARK BIOTECH INC.  
XX PA (CLEV-) CLEVELAND CLINIC FOUND.  
XX PI Feinstein E, Gudkov AV;  
XX WPI; 2003-598393/56.  
XX DR Diagnosing cancer comprises determining the polypeptide or polynucleotide  
XX levels e.g., hepatic lipase, in a sample from a subject, where a higher  
XX level compared to that in a subject free of cancer is indicative of  
XX cancer.  
XX PS Disclosure; SEQ ID NO 311; 272pp; English.  
XX CC The invention relates to a novel method for diagnosing a cancer in a  
XX subject. the method comprises determining, in a sample from the subject,  
XX the level of at least one polypeptide, where a higher level of the  
XX polypeptide compared to the level of the polypeptide in a subject free of  
XX cancer is indicative of cancer. The polypeptide is selected from any of  
XX the polypeptides encoded by the polynucleotides listed in the  
XX specification and polypeptides which are at least 70% homologous to the  
XX polypeptides. The method of the invention has cytostatic activity, and  
XX may have a use in gene therapy. The method is useful in identifying  
XX markers specific for one or several types of cancer, depending on the  
XX tissue origin, which may be used in numerous diagnostic and prognostic  
XX applications as well as cancer type-specific targets for therapeutic  
XX intervention. The compounds that modulate the activity of a tumour  
XX suppressor gene are useful in the treatment of cancer or as anti-cancer  
XX drugs. The present sequence represents a polynucleotide of the invention.  
XX SQ Sequence 2448 BP; 573 A; 728 C; 609 G; 522 T; 0 U; 16 Other;  
Query Match 3.8%; Score 19; DB 9; Length 2448;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 418 GCCCGCGCGCGCGACGACG 436  
DB 315 GCCCGCGCGCGCGACGACG 333  
RESULT 33  
AAX86614/c  
ID AAX86614 standard; cDNA; 2478 BP.

```
XX AAX86614;
XX 15-OCT-1999 (first entry)
XX cDNA encoding an acetylcholinesterase protein/scFv fusion protein.
XX
XX Acetylcholinesterase; AChE; fusion protein; ligand receptor; monomer;
XX ligand detection; marker enzyme; ss.
XX
XX Synthetic.
XX Bungarus fasciatus.
XX
XX Key Location/Qualifiers
XX CDS 1..2478
XX /tag= a
XX /transl_except= (pos: 52..54, aa: Leu)
XX /transl_except= (pos: 55..57, aa: Cys)
XX /transl_except= (pos: 58..60, aa: Ile)
XX /transl_except= (pos: 61..63, aa: Pro)
XX /transl_except= (pos: 64..66, aa: Ser)
XX /transl_except= (pos: 67..69, aa: Cys)
XX /transl_except= (pos: 70..72, aa: Val)
XX /transl_except= (pos: 73..75, aa: Ala)
XX /transl_except= (pos: 76..78, aa: Val)
XX /transl_except= (pos: 79..81, aa: Leu)
XX /transl_except= (pos: 82..84, aa: Gln)
XX /transl_except= (pos: 85..87, aa: Leu)
XX /note= "no termination codon given"
XX
XX FR2773802-A1.
XX
XX 23-JUL-1999.
XX
XX 22-JAN-1998; 98FR-00000656.
XX
XX 22-JAN-1998; 98FR-00000656.
XX
XX (INSP ) INST PASTEUR.
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Bon C, Choumet V, Cousin X;
XX
XX WPI: 1999-471239/40.
XX P-PSDB; AAY30101.
XX
XX A fusion protein comprising an acetyl cholinesterase and ligand receptor,
XX useful for detection of ligands.
XX
XX Claim 20; Page 99-102; 114pp; French.
XX
XX The present sequence encodes an acetylcholinesterase protein (AChE)/ scFv
XX fusion protein of the invention. The specification describes a fusion
XX protein comprising an AChE monomer and a specific ligand receptor. The
XX AChE fusion protein is useful for the production of an AChE monomer in a
XX soluble format. The AChE fusion polypeptide is useful for detection of
XX ligands in samples. AChE is used as a marker enzyme, in a similar manner
XX to peroxidase, alkaline phosphatase and beta-galactosidase. By having
XX AChE fused to a receptor protein, various ligands can be detected by
XX their binding to the receptor portion of the fusion polypeptide
XX
XX Sequence 2478 BP; 494 A; 796 C; 710 G; 478 T; 0 U; 0 Other;
XX
XX Query Match 3.8%; Score 19; DB 2; Length 2478;
XX Best Local Similarity 100.0%; Pred.No. 58;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 407 CCACCGAGCGCGCCCGC 425
XX
XX 712 CCACCGAGCGCGCCCGC 694
XX
XX RESULT 34
```

```
ABQ88107
ID ABQ88107 standard; cDNA; 3081 BP.
XX
XX AC ABQ88107;
XX
XX DT 18-SEP-2002 (first entry)
XX
XX DE Human osteoblast differentiation related cDNA SEQ ID NO 14.
XX
XX KW Human; osteoblast; stem cell differentiation; bone tissue deposition;
XX osteoporosis; osteopathic; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200250301-A2.
XX
XX PD 27-JUN-2002.
XX
XX PF 18-DRC-2001; 2001WO-US048276.
XX
XX PR 18-DRC-2000; 2000US-0255882P.
XX 24-APR-2001; 2001US-0285691P.
XX
XX (GENE-) GENE LOGIC INC.
XX (PROC ) PROCTER & GAMBLE CO.
XX
XX Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
XX Mertz L;
XX
XX WPI: 2002-557663/59.
XX
XX Use of genes and their expression profiles associated with osteoblast
XX differentiation for screening modulators bone formation, for diagnosing
XX or treating e.g. osteoporosis, or as markers for the differentiation
XX process.
XX
XX Claim 1; SEQ ID NO 14; 78pp + Sequence Listing; English.
XX
XX The invention relates to genes and their expression profiles are used
XX for: (a) screening modulators of precursor stem cell differentiation into
XX osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
XX deposition of bone tissue, abnormal rate of osteoblast formation or
XX osteoporosis; or (c) treating or monitoring of bone tissue deposition.
XX cited in (b), or monitoring the progression of bone tissue deposition.
XX Specific conditions include postmenopausal osteoporosis, glucocorticoid
XX osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
XX induced abnormalities in bone formation or bone loss, conditions that
XX involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
XX skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
XX or fibrous dysplasia. The present sequence is that of an osteoblast
XX differentiation associated cDNA marker of the invention. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3081 BP; 613 A; 889 C; 882 G; 697 T; 0 U; 0 Other;
XX
XX Query Match 3.8%; Score 19; DB 6; Length 3081;
XX Best Local Similarity 100.0%; Pred.No. 56;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 12 CGCGAGCAGAACCGCATC 30
XX
XX 2915 CGCGAGCAGAACCGCATC 2933
XX
XX RESULT 35
XX AAI59458
XX ID AAI59458 standard; cDNA; 5092 BP.
XX
XX AAI59458;
XX
XX 22-OCT-2001 (first entry)
```



DE Human polynucleotide SEQ ID NO 1661.  
 DE Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 DE peripheral nervous system; neuropathy; central nervous system; CNS;  
 DE Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 DE amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 DE chemokine; thrombolytic; drug screening; arthritis; inflammation;  
 DE leukaemia; ss.  
 OS Homo sapiens.  
 PN WO200153312-A1.  
 PD 26-JUL-2001.  
 XX 26-DEC-2000; 2000WO-US034263.  
 XX 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 20-JUN-2000; 2000US-00598042.  
 PR 19-JUL-2000; 2000US-00620312.  
 PR 03-AUG-2000; 2000US-00653450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX (HYSEQ-) HYSEQ INC.  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 XX Zhou F, Goodrich R, Dzmanac RI;  
 DR WPI; 2001-442253/47.  
 DR P-PSDB; AAM40302.  
 XX Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX Claim 1; SEQ ID NO 1661; 10078pp; English.  
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
 CC encoded polypeptides (AAM38642-AAM42213) with nontropic, and the  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification  
 XX  
 SQ Sequence 5092 BP; 1310 A; 1441 C; 1448 G; 893 T; 0 U; 0 Other;  
 Query Match 3.8%; Score 19; DB 4; Length 5092;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 27 CATCTGCTGAAGCTGTC 45  
 Db 3813 CATCTGCTGAAGCTGTC 3831  
 RESULT 36  
 ABL62902  
 ID ABL62902 standard; DNA; 5434 BP.  
 XX

AC ABL62902;  
 XX 15-MAY-2002 (first entry)  
 DT Breast cancer related gene sequence SEQ ID NO:1239.  
 DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 DE stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 DE cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
 DE gene; ds.  
 OS Homo sapiens.  
 PN WO200194629-A2.  
 PD 13-DEC-2001.  
 XX 30-MAY-2001; 2001WO-US010838.  
 XX 05-JUN-2000; 2000US-0209473P.  
 PR 05-JUN-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-0233133P.  
 PR 18-SEP-2000; 2000US-0233617P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.  
 PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.  
 PR 25-SEP-2000; 2000US-0235280P.  
 PR 26-SEP-2000; 2000US-0235637P.  
 PR 26-SEP-2000; 2000US-0235638P.  
 PR 27-SEP-2000; 2000US-0235711P.  
 PR 27-SEP-2000; 2000US-0235720P.  
 PR 27-SEP-2000; 2000US-0235840P.  
 PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX (AVAL-) AVALON PHARM.  
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 PI WPI; 2002-188264/24.  
 DR Screening for anti-neoplastic agent involves exposing cells to a chemical  
 XX agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.

PS Claim 1; SEQ ID NO 1239; 44pp; English.  
 XX The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences given in ABL61664 to ABL70110, or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms' tumour  
 SQ Sequence 5434 BP; 1418 A; 1488 C; 1511 G; 1017 T; 0 U; 0 Other;  
 Query Match 3.8%; Score 19; DB 6; Length 5434;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 27 CATCTGCTGAAGCTGGTC 45  
 Db 3813 CATCTGCTGAAGCTGGTC 3831  
 RESULT 37  
 ABL63111  
 ID ABL63111 standard; DNA; 5434 BP.  
 XX ABL63111;  
 XX 15-MAY-2002 (first entry)  
 DT Breast cancer related gene sequence SEQ ID NO:1448.  
 DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;  
 KW gene; ds.  
 XX Homo sapiens.  
 OS WO200194629-A2.  
 PN 13-DEC-2001.  
 PD 30-MAY-2001; 2001WO-US010838.  
 PF 05-JUN-2000; 2000US-0209473P.  
 PR 05-JUN-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-0233133P.  
 PR 18-SEP-2000; 2000US-0233617P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.  
 PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.  
 PR 26-SEP-2000; 2000US-0235280P.  
 PR 26-SEP-2000; 2000US-0235637P.  
 PR 27-SEP-2000; 2000US-0235638P.  
 PR 27-SEP-2000; 2000US-0235711P.  
 PR 27-SEP-2000; 2000US-0235720P.

PR 27-SEP-2000; 2000US-0235840P.  
 PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 28-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237395P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX (AVAL-) AVALON PHARM.  
 PA Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX WPI; 2002-188264/24.  
 DR Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene set.  
 PT Claim 1; SEQ ID NO 1448; 44pp; English.  
 PS The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms' tumour  
 SQ Sequence 5434 BP; 1418 A; 1488 C; 1511 G; 1017 T; 0 U; 0 Other;  
 Query Match 3.8%; Score 19; DB 6; Length 5434;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 27 CATCTGCTGAAGCTGGTC 45  
 Db 3813 CATCTGCTGAAGCTGGTC 3831  
 RESULT 38  
 ABL64407  
 ID ABL64407 standard; DNA; 5434 BP.  
 XX ABL64407;  
 AC 15-MAY-2002 (first entry)  
 DT

XX DE Stomach cancer related gene sequence SEQ ID NO:2744.  
 XX DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
 KW gene; ds.  
 XX OS Homo sapiens.  
 XX PN WO200194629-A2.  
 XX PD 13-DEC-2001.  
 XX PF 30-MAY-2001; 2001WO-US010838.  
 XX PR 05-JUN-2000; 2000US-0209473P.  
 XX PR 05-JUN-2000; 2000US-0209531P.  
 XX PR 18-SEP-2000; 2000US-0233133P.  
 XX PR 18-SEP-2000; 2000US-0233617P.  
 XX PR 20-SEP-2000; 2000US-0234009P.  
 XX PR 20-SEP-2000; 2000US-0234034P.  
 XX PR 20-SEP-2000; 2000US-0234052P.  
 XX PR 22-SEP-2000; 2000US-0234509P.  
 XX PR 22-SEP-2000; 2000US-0234567P.  
 XX PR 25-SEP-2000; 2000US-0234923P.  
 XX PR 25-SEP-2000; 2000US-0234924P.  
 XX PR 25-SEP-2000; 2000US-0235077P.  
 XX PR 25-SEP-2000; 2000US-0235082P.  
 XX PR 25-SEP-2000; 2000US-0235134P.  
 XX PR 25-SEP-2000; 2000US-0235280P.  
 XX PR 26-SEP-2000; 2000US-0235637P.  
 XX PR 26-SEP-2000; 2000US-0235638P.  
 XX PR 27-SEP-2000; 2000US-0235711P.  
 XX PR 27-SEP-2000; 2000US-0235720P.  
 XX PR 27-SEP-2000; 2000US-0235840P.  
 XX PR 27-SEP-2000; 2000US-0235863P.  
 XX PR 28-SEP-2000; 2000US-0236028P.  
 XX PR 28-SEP-2000; 2000US-0236032P.  
 XX PR 28-SEP-2000; 2000US-0236033P.  
 XX PR 28-SEP-2000; 2000US-0236034P.  
 XX PR 28-SEP-2000; 2000US-0236109P.  
 XX PR 28-SEP-2000; 2000US-0236111P.  
 XX PR 29-SEP-2000; 2000US-0236842P.  
 XX PR 29-SEP-2000; 2000US-0236891P.  
 XX PR 02-OCT-2000; 2000US-0237172P.  
 XX PR 02-OCT-2000; 2000US-0237173P.  
 XX PR 02-OCT-2000; 2000US-0237278P.  
 XX PR 02-OCT-2000; 2000US-0237294P.  
 XX PR 02-OCT-2000; 2000US-0237295P.  
 XX PR 02-OCT-2000; 2000US-0237316P.  
 XX PR 03-OCT-2000; 2000US-0237425P.  
 XX PR 03-OCT-2000; 2000US-0237598P.  
 XX PR 03-OCT-2000; 2000US-0237604P.  
 XX PR 03-OCT-2000; 2000US-0237606P.  
 XX PR 03-OCT-2000; 2000US-0237608P.  
 XX PR 01-NOV-2000; 2000US-0244867P.  
 XX PR 01-NOV-2000; 2000US-0245084P.  
 XX PA (AVAL-) AVALON PHARM.  
 XX PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX WPI; 2002-188264/24.  
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.  
 XX Claim 1; SEQ ID NO 2744; 44pp; English.  
 XX The present invention describes a method (M1) for screening for an anti-

CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABU61664  
 CC to ABU70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's  
 CC tumour  
 XX Sequence 5434 BP; 1418 A; 1488 C; 1511 G; 1017 T; 0 U; 0 Other;  
 SQ Best Local Similarity 3.8%; Score 19; DB 6; Length 5434;  
 Query Match 100.0%; Pred. No. 52;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 27 CATCTGCTGAAGCTGTC 45  
 DB 3813 CATCTGCTGAAGCTGTC 3831  
 RESULT 39  
 ID APT33842 standard; cDNA; 5434 BP.  
 XX AC APT33842;  
 XX DT 29-MAY-2003 (first entry)  
 XX DNMT1 cDNA sequence SEQ ID No 1.  
 XX Cytostatic; methyl transferase inhibitor; DNA methyl transferase isoform;  
 XX gene therapy; anti-DNA methyl transferase oligonucleotide; inhibitor;  
 XX cell proliferation; neoplasia; DNMT 1; enzyme; ss.  
 XX Unidentified.  
 XX WO200291926-A2.  
 XX 21-NOV-2002.  
 XX 13-MAY-2002; 2002WO-IB003120.  
 XX 11-MAY-2001; 2001US-0290202P.  
 XX 11-MAY-2001; 2001US-0290212P.  
 XX (METH-) METHYLGENE INC.  
 XX Macleod AR;  
 XX WPI; 2003-148369/14.  
 XX New inhibitors of DNA methyl transferase isoforms, e.g. anti-DNA methyl  
 XX transferase oligonucleotides or small molecule inhibitors of DNA methyl  
 XX transferase, useful for treating cell proliferative and differentiation  
 XX disorders.  
 XX Claim 5; Fig 1B; 76pp; English.  
 XX The invention relates to an agent that inhibits one or more specific DNA  
 CC methyl transferase isoforms (but not all DNA methyl transferase  
 CC isoforms), such as an anti-DNA methyl transferase oligonucleotide or a  
 CC small molecule inhibitor of DNA methyl transferase. The agents,  
 CC oligonucleotides, inhibitors and methods are useful for identifying  
 CC specific inhibition of specific DNA methyl transferase isoforms involved  
 CC in cell proliferation and/or differentiation, and thus providing a

CC treatment for cell proliferative and/or differentiation disorders, e.g.  
CC neoplasia. This polynucleotide sequence represents a DNMT1 cDNA sequence  
CC relating to the invention  
XX  
SQ Sequence 5434 BP; 1418 A; 1488 C; 1511 G; 1017 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 7; Length 5434;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 27 CATCTGCTGAAGCTGGTC 45  
DB 3813 CATCTGCTGAAGCTGGTC 3831

RESULT 40  
AAS78583  
ID AAS78583 standard; cDNA; 5794 BP.  
XX  
AC AAS78583;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #14387.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
XN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Dmanac RT, Liu C, Tang YT;  
XX  
XN WPI; 2001-639362/73.  
DR P-PSDB; ABG14396.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 14387; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have application in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5794 BP; 1466 A; 1580 C; 1617 G; 1128 T; 0 U; 3 Other;  
Query Match 3.8%; Score 19; DB 5; Length 5794;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 27 CATCTGCTGAAGCTGGTC 45  
DB 3829 CATCTGCTGAAGCTGGTC 3847

RESULT 41  
AAV99129  
ID AAV99129 standard; DNA; 19440 BP.  
XX  
AC AAV99129;  
XX  
DT 16-MAR-1999 (first entry)  
XX  
DE DNA methyltransferase genomic sequence.  
XX  
KW DNA methyltransferase; DNA Mefase; antisense oligonucleotide; human;  
KW cellular growth; tumour growth inhibition; silenced gene activation;  
KW beta thalassemia; sickle cell anemia; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 115  
FT /tag= a  
FT /note= "this nucleotide is not present in AAV99092"  
FT 6741  
FT /tag= b  
FT /note= "this nucleotide is not present in AAV99094"  
FT 6744..6745  
FT /tag= c  
FT /note= "approximately 600 unspecified nucleotides are  
FT present between these bases"  
FT 7097..7098  
FT /tag= d  
FT /note= "approximately 1000 unspecified nucleotides are  
FT present between these bases"  
FT 7556..7557  
FT /tag= e  
FT /note= "approximately 1500 unspecified nucleotides are  
FT present between these bases"  
FT 7866..7867  
FT /tag= f  
FT /note= "approximately 700 unspecified nucleotides are  
FT present between these bases"  
FT 8542..8543  
FT /tag= g  
FT /note= "approximately 1000 unspecified nucleotides are  
FT present between these bases"  
FT 8807..8808  
FT /tag= h  
FT /note= "1700 unspecified nucleotides are present between  
FT these bases"  
FT 9099..9100  
FT /tag= i  
FT /note= "570 unspecified nucleotides are present between  
FT these bases"  
FT 9106..9108  
FT /tag= j  
FT /note= "these nucleotides are not present in AAV99103"  
FT 9355..9356  
FT /tag= k  
FT /note= "350 unspecified nucleotides are present between  
FT these bases"  
FT 9662..9663  
FT /tag= l

```
FT /note= "300 unspecified nucleotides are present between
FT these bases"
FT /tag= m
FT /note= "approximately 450 unspecified nucleotides are
FT present between these bases"
FT /tag= n
FT /note= "approximately 900 unspecified nucleotides are
FT present between these bases"
FT /tag= o
FT /note= "approximately 180 unspecified nucleotides are
FT present between these bases"
FT /tag= p
FT /note= "these nucleotides as stated as being from
FT AAV99111 in the figure, but are actually from AAV99112"
FT /tag= q
FT /note= "approximately 1500 unspecified nucleotides are
FT present between these bases"
FT /tag= r
FT /note= "this nucleotide is not present in AAV99112"
FT /tag= s
FT /note= "approximately 150 unspecified nucleotides are
FT present between these bases"
FT /tag= t
FT /note= "400 unspecified nucleotides are present between
FT these bases"
FT /tag= u
FT /note= "approximately 850 unspecified nucleotides are
FT present between these bases"
FT /tag= v
FT /note= "approximately 450 unspecified nucleotides are
FT present between these bases"
FT /tag= w
FT /note= "approximately 520 unspecified nucleotides are
FT present between these bases"
FT /tag= x
FT /note= "approximately 950 unspecified nucleotides are
FT present between these bases"
FT /tag= y
FT /note= "approximately 500 unspecified nucleotides are
FT present between these bases"
FT /tag= z
FT /note= "approximately 250 unspecified nucleotides are
FT present between these bases"
FT /tag= aa
FT /note= "approximately 1200 unspecified nucleotides are
FT present between these bases"
FT /tag= ab
FT /note= "approximately 600 unspecified nucleotides are
FT present between these bases"
FT /tag= ac
FT /note= "approximately 600 unspecified nucleotides are
FT present between these bases"
FT /tag= ad
FT /note= "approximately 800 unspecified nucleotides are
FT present between these bases"
```

```
FT misc_feature 18254. .18253
FT /tag= ae
FT /note= "approximately 400 unspecified nucleotides are
FT present between these bases"
FT /tag= af
FT /note= "approximately 1450 unspecified nucleotides are
FT present between these bases"
FT /tag= ag
FT /note= "approximately 550 unspecified nucleotides are
FT present between these bases"
FT WO9854313-A2.
FT 03-DEC-1998.
FT 29-MAY-1998; 98WO-IB001107.
FT 30-MAY-1997; 97US-00866340.
FT 17-DEC-1997; 97US-0069865P.
FT (UYMC-) UNIV MCGILL.
FT Szyf M, Bigey P, Ramchandani S;
FT WPI; 1999-059833/05.
FT New DNA methyltransferase nucleotide sequences - used particularly to
FT develop antisense oligonucleotides for diagnostic and therapeutic
FT purposes, particularly for inhibiting tumour growth.
FT Claim 1; Fig 1; 108pp; English.
FT
FT The present sequence represents a full length DNA methyltransferase (DNA
FT Metase) genomic sequence. It is derived from fragments AAV99091-128. The
FT sequences can be used to derive antisense oligonucleotides which inhibit
FT DNA Metase expression. The antisense oligonucleotides can be used in
FT investigating the role of DNA Metase in cellular growth. They can be
FT administered at different points in the cell cycle, or in conjugation
FT with promoters or inhibitors of cell growth to determine the role of DNA
FT Metase in the growth of the cell type of interest. The antisense
FT oligonucleotides can also be used for inhibiting tumour growth in a
FT mammal, or to activate silenced genes to provide a missing gene function.
FT This ameliorates disease symptoms, e.g. in beta thalassemia and sickle
FT cell anemia. The antisense oligonucleotides can also be used as an
FT analytical and diagnostic tools and a potentiators of transgenic plant
FT and animal studies
FT
FT SQ Sequence 19440 BP; 4471 A; 4929 C; 4936 G; 4915 T; 0 U; 189 Other;
FT
FT Query Match 3.8%; Score 19; DB 2; Length 19440;
FT Best Local Similarity 100.0%; Pred.No.45;
FT Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT QY 27 CATCCTGCTGAAGCTGTC 45
FT |||||
FT Db 16076 CATCCTGCTGAAGCTGTC 16094
FT
FT RESULT 42
FT AAD54645
FT ID AAD54645 standard; DNA; 113193 BP.
FT AC AAD54645;
FT XX
FT DT 26-JUN-2003 (first entry)
FT XX Streptomyces nodosus amphotericin (amph) biosynthetic gene cluster.
FT DE Polyene; antibiotic; amphotericin; amph; polyketide; enzyme; ds.
FT XX Streptomyces nodosus.
FT OS
```

```

XX FH Location/Qualifiers
XX FT complement(4. .1824)
XX FT /*tag= a
XX FT /product= "ABC transporter encoded by S. nodosus amphG
XX FT gene"
XX FT complement(1805. .3628)
XX FT /*tag= b
XX FT /product= "ABC transporter encoded by S. nodosus amphH
XX FT gene"
XX FT 3840. .4874
XX FT /*tag= c
XX FT /product= "GDP-mannose dehydratase encoded by S. nodosus
XX FT amphDII gene"
XX FT 5042. .33574
XX FT /*tag= d
XX FT /product= "Polyketide synthase multienzyme housing
XX FT extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
XX FT 33584. .50518
XX FT /product= "Polyketide synthase multienzyme housing
XX FT extension modules 15, 16 and 17 encoded by S. nodosus
XX FT amphi gene"
XX FT 50571. .56675
XX FT /*tag= f
XX FT /product= "Polyketide synthase multienzyme housing
XX FT extension modules 18 and thioesterase encoded by S.
XX FT nodosus amphi gene"
XX FT 56823. .58019
XX FT /*tag= g
XX FT /product= "Cytochrome P450 encoded by S. nodosus amphi
XX FT gene"
XX FT 58139. .58648
XX FT /*tag= h
XX FT /product= "ORF1, hypothetical protein"
XX FT complement(58756. .59610)
XX FT /*tag= i
XX FT /product= "ORF2, hypothetical protein"
XX FT 59869. .61470
XX FT /*tag= j
XX FT /product= "ORF3, hypothetical protein"
XX FT /transl_except= (pos:59869. .59871, aa:Met)
XX FT complement(61798. .61995)
XX FT /*tag= k
XX FT /product= "Ferredoxin encoded by S. nodosus amphi gene"
XX FT complement(62051. .63250)
XX FT /*tag= l
XX FT /product= "Cytochrome P450 encoded by S. nodosus amphi
XX FT gene"
XX FT complement(63250. .64308)
XX FT /*tag= m
XX FT /product= "NDP-sugar aminotransferase encoded by S.
XX FT nodosus amphi gene"
XX FT complement(64324. .65775)
XX FT /*tag= n
XX FT /product= "Glycosyl transferase encoded by S. nodosus
XX FT amphi gene"
XX FT /transl_except= (pos:65773. .65775, aa:Met)
XX FT 66081. .70319
XX FT /*tag= o
XX FT /product= "Polyketide synthase multienzyme housing
XX FT loading module encoded by S. nodosus amphi gene"
XX FT 70366. .79938
XX FT /*tag= p
XX FT /product= "Polyketide synthase multienzyme housing
XX FT extension modules 1 and 2 encoded by S. nodosus by amphi
XX FT gene"
XX FT 79956. .112709
XX FT /*tag= q
XX FT /product= "Polyketide synthase multienzyme housing
XX FT extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
XX FT nodosus by amphi gene"

```

```

XX PN WO200297082-A2.
XX PD 05-DEC-2002.
XX PF 27-MAY-2002; 2002WO-IE000071.
XX PF 31-MAY-2001; 2001IE-00000527.
XX PR (UYDU-) UNIV COLLEGE DUBLIN.
XX PA Caffrey JP;
XX PI WPI; 2003-201271/19.
XX DR P-PSDB; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
XX DR AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
XX DR AAE36129, AAE36130, AAE36131, AAE36132.
XX FT Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
XX FT for preparing amphotericin derivative or analog antibiotic agent with
XX FT altered properties, in biosynthesis of polyketide other than
XX FT amphotericin.
XX PS Claim 1; Page 52-114; 276pp; English.
XX CC The invention relates to the gene cluster encoding the polypeptides
XX CC responsible for the biosynthesis of the polyene antibiotic amphotericin
XX CC (amph) of Streptomyces nodosus. Polynucleotides of the invention are
XX CC useful for preparing amphotericin derivatives or analogue antibiotic
XX CC agents with altered properties and in the biosynthesis of polyketides
XX CC other than amphotericin. amphDII, amphDII or amphDI mutants are useful
XX CC for producing amphotericin derivatives glycosylated with alternative
XX CC sugars; amphDII or amphDII gene sequences are useful in engineered
XX CC biosynthesis of perosaminyl-amphoteronolide B; amphDII or amphDII and
XX CC amphi gene sequences are useful in the engineered biosynthesis of
XX CC perosaminyl-16-desacetoxy-16-methyl amphoteronolide B; amphDII, amphDII
XX CC and amphi gene sequences are useful for preparing polypeptides capable
XX CC of addition of mycosamine to a polyketide other than amphoteronolide A or
XX CC B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
XX CC The present sequence is S. nodosus amphi biosynthetic gene cluster
XX SQ Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;
XX Query Match 3.8%; Score 19; DB 7; Length 113193;
XX Best Local Similarity 100.0%; Pred. No. 37;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 121 GGCTTCGACTCGCTGGCCG 139
DB 69945 GGCTTCGACTCGCTGGCCG 69963
RESULT 43
AAI64291/c
ID AAI64291 standard; DNA; 128139 BP.
XX AC AAI64291;
XX DT 22-APR-2002 (first entry)
XX DE RRV genome nucleotide sequence.
XX RRV; rhesus rhadinovirus; Japanese macaque virus; multiple sclerosis;
XX JMHV; cytostatic; antiasthmatic; antiallergic; dermatological; vulnery;
XX gene therapy; leucopenia; thrombocytopaenia; inflammatory disease;
XX asthma; allergy; dermatitis; virus; ds.
XX OS Macaca mulatta rhadinovirus 17577.
XX FH Key Location/Qualifiers
XX FT CDS 1353. .2674
XX FT /*tag= a
XX FT /product= "RRV RI"

```

```
FT FT CDS complement(2692..3258)
FT FT /tag= b
FT FT /product= "dihydrofolate reductase"
FT FT /label= RVV_ORF2
FT FT /note= "has similarity to Kaposi's sarcoma-associated
FT FT virus (KSHV) open reading frame (ORF) 2"
FT FT 3676..5613
FT FT /tag= c
FT FT /product= "complement binding protein"
FT FT /label= RVV_ORF4
FT FT /note= "has similarity to KSHV ORF4"
FT FT 6045..9443
FT FT /tag= d
FT FT /product= "ssDNA binding protein"
FT FT /label= RVV_ORF6
FT FT /note= "has similarity to KSHV ORF6"
FT FT 9468..11528
FT FT /tag= e
FT FT /product= "transport protein"
FT FT /label= RVV_ORF7
FT FT /note= "has similarity to KSHV ORF7"
FT FT 11515..14004
FT FT /tag= f
FT FT /product= "glycoprotein B"
FT FT /label= RVV_ORF8
FT FT /note= "has similarity to KSHV ORF8"
FT FT 14122..17166
FT FT /tag= g
FT FT /product= "DNA polymerase protein"
FT FT /label= RVV_ORF9
FT FT /note= "has similarity to KSHV ORF9"
FT FT 17261..18511
FT FT /tag= h
FT FT /label= RVV_ORF10
FT FT /note= "has similarity to KSHV ORF10"
FT FT 18520..19749
FT FT /tag= i
FT FT /label= RVV_ORF11
FT FT /note= "has similarity to KSHV ORF11"
FT FT complement(19921..20544)
FT FT /tag= j
FT FT /product= "RVV R2"
FT FT /label= RVV_ORF
FT FT /note= "has similarity to KSHV interleukin (IL)-6 gene"
FT FT complement(20777..21778)
FT FT /tag= k
FT FT /product= "thymidylate synthase"
FT FT /label= RVV_ORF70
FT FT /note= "has similarity to KSHV ORF70"
FT FT complement(22245..22592)
FT FT /tag= l
FT FT /product= "RVV R3"
FT FT /note= "has similarity to KSHV K4 viral MIP gene"
FT FT 26846..27409
FT FT /tag= m
FT FT /product= "Bcl2-homologue"
FT FT /label= RVV_ORF16
FT FT /note= "has similarity to KSHV ORF16"
FT FT complement(27515..29125)
FT FT /tag= n
FT FT /label= RVV_ORF17
FT FT /note= "has similarity to KSHV ORF17"
FT FT 28998..29897
FT FT /tag= o
FT FT /label= RVV_ORF18
FT FT /note= "has similarity to KSHV ORF18"
FT FT complement(29905..31548)
FT FT /tag= p
FT FT /product= "tegument protein"
FT FT /label= RVV_ORF19
FT FT /note= "has similarity to KSHV ORF19"
FT FT complement(31043..32095)
FT FT /tag= q

FT FT CDS /label= RVV_ORF20
FT FT /note= "has similarity to KSHV ORF20"
FT FT 32094..33767
FT FT /tag= r
FT FT /product= "thymidine kinase"
FT FT /label= RVV_ORF21
FT FT /note= "has similarity to KSHV ORF21"
FT FT 33754..35868
FT FT /tag= s
FT FT /product= "glycoprotein H"
FT FT /label= RVV_ORF22
FT FT /note= "has similarity to KSHV ORF22"
FT FT complement(35865..37073)
FT FT /tag= t
FT FT /label= RVV_ORF23
FT FT /note= "has similarity to KSHV ORF23"
FT FT complement(37123..39321)
FT FT /tag= u
FT FT /label= RVV_ORF24
FT FT /note= "has similarity to KSHV ORF24"
FT FT 39323..43459
FT FT /tag= v
FT FT /product= "major capsid protein"
FT FT /label= RVV_ORF25
FT FT /note= "has similarity to KSHV ORF25"
FT FT 43491..44408
FT FT /tag= w
FT FT /product= "capsid protein"
FT FT /label= RVV_ORF26
FT FT /note= "has similarity to KSHV ORF26"
FT FT 44433..45242
FT FT /tag= x
FT FT /label= RVV_ORF27
FT FT /note= "has similarity to KSHV ORF27"
FT FT 45408..45683
FT FT /tag= y
FT FT /label= RVV_ORF28
FT FT /note= "has similarity to KSHV ORF28"
FT FT complement(45733..46779)
FT FT /tag= z
FT FT /label= RVV_ORF29b
FT FT /note= "has similarity to KSHV ORF29b"
FT FT 46905..47135
FT FT /tag= aa
FT FT /label= RVV_ORF30
FT FT /note= "has similarity to KSHV ORF30"
FT FT 47093..47746
FT FT /tag= ab
FT FT /label= RVV_ORF31
FT FT /note= "has similarity to KSHV ORF31"
FT FT 47683..49077
FT FT /tag= ac
FT FT /label= RVV_ORF32
FT FT /note= "has similarity to KSHV ORF32"
FT FT 49049..50059
FT FT /tag= ad
FT FT /label= RVV_ORF33
FT FT /note= "has similarity to KSHV ORF33"
FT FT complement(49977..50960)
FT FT /tag= ae
FT FT /label= RVV_ORF29a
FT FT /note= "has similarity to KSHV ORF29a"
FT FT 50959..51942
FT FT /tag= af
FT FT /label= RVV_ORF34
FT FT /note= "has similarity to KSHV ORF34"
FT FT 51923..52372
FT FT /tag= ag
FT FT /label= RVV_ORF35
FT FT /note= "has similarity to KSHV ORF35"
FT FT 52278..53585
FT FT /tag= ah
FT FT /product= "kinase"
```



FT /label= RRV\_ORF36  
 FT /note= "has similarity to KSHV ORF36"  
 FT 53566. .55008  
 FT \*tag= ai  
 FT /product= "alkaline exonuclease"  
 FT /label= RRV\_ORF37  
 FT /note= "has similarity to KSHV ORF37"  
 FT 54963. .55172  
 FT \*tag= aj  
 FT /label= RRV\_ORF38  
 FT /note= "has similarity to KSHV ORF38"  
 FT complement(55255. .56391)  
 FT \*tag= ak  
 FT /product= "glycoprotein M"  
 FT /label= RRV\_ORF39  
 FT /note= "has similarity to KSHV ORF39"  
 FT 56526. .57932  
 FT \*tag= al  
 FT /product= "helicase/primase"  
 FT /label= RRV\_ORF40  
 FT /note= "has similarity to KSHV ORF40"  
 FT 57917. .58528  
 FT \*tag= am  
 FT /product= "helicase/primase"  
 FT /label= RRV\_ORF41  
 FT /note= "has similarity to KSHV ORF41"  
 FT complement(58525. .59343)  
 FT \*tag= an  
 FT /label= RRV\_ORF42  
 FT /note= "has similarity to KSHV ORF42"  
 FT complement(59297. .61027)  
 FT \*tag= ao  
 FT /product= "capsid protein"  
 FT /label= RRV\_ORF43  
 FT /note= "has similarity to KSHV ORF43"

Query Match 3.8%; Score 19; DB 6; Length 128139;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 TGATCGACGACCTGGACAT 461  
 |||||  
 Db 53480 TGATCGACGACCTGGACAT 53462

RESULT 44  
 AAC64754/c  
 ID AAC64754 standard; DNA; 133719 BP.  
 XX  
 AC AAC64754;  
 XX  
 DT 28-FEB-2001 (first entry)  
 XX  
 DE Macaca mulatta rhadinovirus 17577 (RRV) genome sequence SEQ ID NO:1.  
 XX  
 KW Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;  
 KW genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;  
 KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;  
 KW cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;  
 KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;  
 KW splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia;  
 KW ds.  
 XX  
 OS Macaca mulatta rhadinovirus 17577.  
 XX  
 PN WO200028040-A2.  
 XX  
 XX 18-MAY-2000.  
 PD  
 XX 05-NOV-1999; 99WO-US026260.  
 PF  
 XX 06-NOV-1998; 98US-0107507P.  
 PR  
 PR 20-NOV-1998; 98US-0109409P.

XX (UYOR-) UNIV OREGON HEALTH SCI.  
 XX PA Wong SW, Axthelm MK, Searles RP;  
 XX PI WPI; 2000-376552/32.  
 XX DR  
 XX PT New rhesus rhadino virus for producing non-human primate model useful for  
 XX testing potential treatments and efficacy of the candidate vaccine for  
 XX conditions associated with RRV infection.  
 XX PT  
 XX PS Claim 2; Page 83-122; 141pp; English.  
 XX  
 CC The present invention describes a novel rhesus macaque rhadinovirus  
 CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the  
 CC RRV genome sequence, and AAB53213 to AAB53204 represent the proteins  
 CC encoded by the genome sequence. The present invention also specifically  
 CC claims the individual open reading frame (ORF) nucleotide sequences from  
 CC the genome which encode the individual proteins, but these sequences are  
 CC not given. A non-human animal infected with RRV can be used for testing  
 CC the efficacy of drug in the treatment of condition associated with  
 CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative  
 CC disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,  
 CC hypergammaglobulinemia or autoimmune haemolytic anaemia, by  
 CC administering the drug to an immuno-compromised non-human primate  
 CC preferably Rhesus macaque monkey obtained by as a result of infection by  
 CC Simian Immunodeficiency Virus (SIV). RRV is useful for producing non-  
 CC human primate model for testing potential treatments for conditions  
 CC associated with RRV infection. It is also useful for testing the efficacy  
 CC of the candidate vaccine against RRV infection or conditions associated  
 CC with its infection by administering the vaccine to the subject capable of  
 CC infection with RRV, inoculating the subject with RRV and observing the  
 CC effect of vaccine. AAC64755 to AAC64765 and AAB53205 to AAB53213  
 CC represent sequence used in the exemplification of the present invention  
 XX  
 SQ Sequence 133719 BP; 32746 A; 35648 C; 34521 G; 30804 T; 0 U; 0 Other;  
 Query Match 3.8%; Score 19; DB 3; Length 133719;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 443 TGATCGACGACCTGGACAT 461  
 |||||  
 Db 59060 TGATCGACGACCTGGACAT 59042  
 RESULT 45  
 ACD93990  
 ID ACD93990 standard; cDNA; 276 BP.  
 XX  
 AC ACD93990;  
 XX  
 DT 23-SEP-2003 (first entry)  
 XX  
 DE Human colon cancer cell expressed cDNA #2402.  
 XX  
 KW Open reading frame detection; genome sequencing; colon cancer;  
 KW breast cancer; population genome analysis; genetic shift; cancer;  
 KW antibiotic resistance; antibiotic non-tolerance; congenital disease;  
 KW agriculture; food crop genome; resistance gene; retrovirus;  
 KW influenza virus; eukaryotic pathogen detection; trypanosome; plasmodium;  
 KW gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002155438-A1.  
 XX  
 XX 24-OCT-2002.  
 PD  
 XX 27-SEP-1999; 99US-00406117.  
 PF  
 XX 20-NOV-1998; 98US-00196716.  
 PR  
 XX

(SIMP/) SIMPSON A J G.  
(NETO/) NETO E D.  
(BREN/) BRENTANI R R.

Simpson AJG, Neto ED, Brentani RR;  
WPI; 2003-182626/18.

Determining open reading frames of genome of an organism e.g. a human  
suffering from cancer involves use of single oligonucleotide primer at  
low stringency for preparing single-stranded cDNA from mRNA of  
individual.

Example 9; Page 364; 959pp; English.

The invention describes a method of determining open reading frames in  
the genome of organism, comprising contacting mRNA from cell of organism  
with a single oligonucleotide primer (I) at low stringency, preparing  
single-stranded cDNA by reverse transcribing mRNA with (I), amplifying  
cDNA, sequencing the product, and repeating the contacting, preparing  
and amplifying steps with different primers and sequencing resulting  
nucleic acids. The method is useful for: determining that a known  
nucleotide sequence from a genome of an organism corresponds to a  
nucleotide sequence of an open reading frame; for preparing a contig,  
nucleic acid molecule from a genome of an organism; and for sequencing  
all or part of a genome of an organism. mRNA is obtained from mammalian  
or human cell which is associated with a pathological condition e.g. a  
colon cancer or breast cancer cell. The method is useful for analyses of  
populations of subjects and can be used to carry out genetic analyses of  
large or small populations. Further, it can be used to study living  
systems to determine if, e.g. there have been genetic shifts which render  
an individual or population more or less likely to be afflicted with  
diseases such as cancer, to determine antibiotic resistance or non-  
tolerance, and so forth. The method can also be used in the study of  
congenital diseases, and the risk of affliction to a foetus, as well as  
the study of whether the conditions are likely to be passed to offspring  
through ova or sperm. The analyses for pathological conditions can be  
carried out in all animals, plants, birds, fish, etc. Using this method,  
in the area of agriculture, for example the genomes of food crops can be  
studied to determine if resistance genes are present, defects in plant  
genomes can also be studied in this way. Similarly, the method permits  
determination of the pathogens which integrate into the genome, such as  
retroviruses and other integrating viruses such as influenza virus, have  
undergone shifts or mutations, which may require different approaches to  
therapy. This method is also applied to eukaryotic pathogens, such as  
trypanosomes, different types of Plasmodium, etc. The method essentially  
eliminates sequencing of non-coding portions. This sequence represents a  
polynucleotide isolated from human colon cancer cell cDNA library

SQ Sequence 276 BP; 41 A; 83 C; 88 G; 64 T; 0 U; 0 Other;

Query Match 3.6%; Score 18; DB 7; Length 276;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 CACAGCGGCGCGAAGGC 90

Db 228 CACAGCGGCGCGAAGGC 245

Search completed: June 16, 2004, 17:43:44  
Job time : 253 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:48:58 ; Search time 57.3333 Seconds  
(without alignments)  
4849.367 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543

Perfect score: 501

Sequence: 1 gcgggtcccccagccgagca.....ccctcgggcagagctga 501

Scoring table: OLIGO NUC

Gapop\_60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 15

Total number of hits satisfying chosen parameters: 778

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgm2\_6/ptodata/2/ina/5A COMB.seq:\*  
2: /cgm2\_6/ptodata/2/ina/5B COMB.seq:\*  
3: /cgm2\_6/ptodata/2/ina/6A COMB.seq:\*  
4: /cgm2\_6/ptodata/2/ina/6B COMB.seq:\*  
5: /cgm2\_6/ptodata/2/ina/PCTUS COMB.seq:\*  
6: /cgm2\_6/ptodata/2/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	5.0	44377	2	US-08-804-227C-7
2	25	5.0	44377	2	US-08-804-198-1
3	23	4.6	748	3	US-09-154-083-1
4	23	4.6	753	3	US-09-154-083-28
5	23	4.6	4403765	3	US-09-103-840A-2
6	23	4.6	4411529	3	US-09-103-840A-1
7	21	4.2	47981	4	US-09-679-279-1
8	20	4.0	1410	4	US-09-252-991A-10147
9	20	4.0	1623	4	US-09-252-991A-9839
10	20	4.0	4026	4	US-09-252-991A-10214
11	20	4.0	20235	1	US-07-642-734C-3
12	20	4.0	20235	3	US-08-439-009A-3
13	20	4.0	43280	2	US-08-804-227C-1
14	19	3.8	375	3	US-08-981-189B-16
15	19	3.8	451	3	US-08-866-340-24
16	19	3.8	452	3	US-09-103-875-30
17	19	3.8	953	4	US-09-452-239-45
18	19	3.8	1049	4	US-09-452-239-43
19	19	3.8	1368	1	US-08-578-590-1
20	19	3.8	1497	4	US-09-252-991A-2256
21	19	3.8	1950	4	US-09-252-991A-2425
22	18	3.6	579	4	US-09-252-991A-9199
23	18	3.6	717	4	US-09-252-991A-8972
24	18	3.6	819	4	US-09-252-991A-8861
25	18	3.6	1656	4	US-09-252-991A-8752
26	18	3.6	1791	1	US-08-399-646-3
27	18	3.6	1791	1	US-08-607-321-3

28	18	3.6	1791	2	US-08-961-240-3
29	18	3.6	1791	2	US-08-605-501-3
30	18	3.6	2056	1	US-08-393-646-13
31	18	3.6	2056	1	US-08-607-321-13
32	18	3.6	2056	2	US-08-961-240-13
33	18	3.6	2056	2	US-08-605-501-13
34	18	3.6	2508	4	US-09-252-991A-793
35	18	3.6	2577	4	US-09-252-991A-828
36	18	3.6	49377	1	US-08-764-233A-1
37	17	3.4	408	4	US-09-252-991A-2185
38	17	3.4	411	4	US-09-252-991A-8838
39	17	3.4	477	4	US-09-252-991A-12519
40	17	3.4	546	4	US-09-252-991A-5395
41	17	3.4	546	4	US-09-252-991A-12682
42	17	3.4	561	3	US-09-154-083-13
43	17	3.4	617	3	US-09-146-221-9
44	17	3.4	723	4	US-09-252-991A-5281
45	17	3.4	744	4	US-09-252-991A-5345

## ALIGNMENTS

RESULT 1  
US-08-804-227C-7  
; Sequence 7, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 350..14002  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14046..20036  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20110..31284  
; FEATURE:

```
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
US-08-804-227C-7

Query Match      5.0%; Score 25; DB 2; Length 44377;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 TCGGCTTCGACTCGCTGCGCGCGGT 143
Db 35710 TCGGCTTCGACTCGCTGCGCGCGGT 35734

RESULT 2
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rosteck, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-198-1

; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
US-08-804-227C-7

Query Match      5.0%; Score 25; DB 2; Length 44377;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 TCGGCTTCGACTCGCTGCGCGCGGT 143
Db 35710 TCGGCTTCGACTCGCTGCGCGCGGT 35734

RESULT 3
US-09-154-083-1/c
; Sequence 1, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; TITLE OF INVENTION: Constructs Therefor
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 748
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-09-154-083-1

Query Match      4.6%; Score 23; DB 3; Length 748;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GCCACCGGGCTGCGGCTGCCCGC 191
Db 366 GCCACCGGGCTGCGGCTGCCCGC 344

RESULT 4
US-09-154-083-28/c
; Sequence 28, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; TITLE OF INVENTION: Constructs Therefor
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-09-154-083-28

Query Match      4.6%; Score 23; DB 3; Length 753;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GCCACCGGGCTGCGGCTGCCCGC 191
Db 371 GCCACCGGGCTGCGGCTGCCCGC 349

RESULT 5
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
```

APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2  
Query Match 4.6%; Score 23; DB 3; Length 4403765;  
Best Local Similarity 100.0%; Pred.No. 0.051;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 193 ACGCTGATCTTCGACTACCCAC 215  
Db 1872270 ACGCTGATCTTCGACTACCCAC 1872292  
RESULT 6  
US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1  
Query Match 4.6%; Score 23; DB 3; Length 4411529;  
Best Local Similarity 100.0%; Pred.No. 0.051;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 193 ACGCTGATCTTCGACTACCCAC 215  
Db 1881395 ACGCTGATCTTCGACTACCCAC 1881417  
RESULT 7  
US-09-679-279-1  
Sequence 1, Application US/09679279  
Patent No. 6524841  
GENERAL INFORMATION:  
APPLICANT: McDaniel, Robert  
APPLICANT: Volchegursky, Yanina  
TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic  
TITLE OF INVENTION: Genes and Uses Thereof  
FILE REFERENCE: 300622004700  
CURRENT APPLICATION NUMBER: US/09/679,279  
CURRENT FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/158,305

PRIOR FILING DATE: 1999-10-08  
PRIOR APPLICATION NUMBER: US 60/190,024  
PRIOR FILING DATE: 2000-03-17  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 47981  
TYPE: DNA  
ORGANISM: Micromonospora megalomicea  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(144)  
OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;  
OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (928)...(2061)  
OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;  
OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (2072)...(3382)  
OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCIII homolog),  
OTHER INFORMATION: TDP-megosamine glycosyltransferase;  
OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (3462)...(4634)  
OTHER INFORMATION: megG (megY), mycarosyl acyltransferase, mycarose O-acyltransferase  
OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (4651)...(5775)  
OTHER INFORMATION: megDII, deoxysugar transaminase (eryCI, DmUV homolog),  
OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;  
OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (5822)...(6595)  
OTHER INFORMATION: megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);  
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (6592)...(7197)  
OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dmUV hc  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;  
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (7220)...(8206)  
OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dmV homolog),  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (8228)...(9220)  
OTHER INFORMATION: megBII-1 (megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;  
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (9226)...(10479)  
OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;  
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (10483)...(11424)  
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (12181)...(22821)  
OTHER INFORMATION: megAI, SEQ ID NO: 13= translated amino acid sequence  
NAME/KEY: misc feature  
LOCATION: (12505)...(13470)  
OTHER INFORMATION: megAL, AT-L  
NAME/KEY: misc feature  
LOCATION: (13576)...(13791)  
OTHER INFORMATION: megAI, ACP-L  
NAME/KEY: misc feature  
LOCATION: (13849)...(15126)  
OTHER INFORMATION: megAL, KSI  
NAME/KEY: misc feature

```

; LOCATION: (15427)...(16476)
; OTHER INFORMATION: megAI, AT1
; NAME/KEY: misc feature
; LOCATION: (17155)...(17694)
; OTHER INFORMATION: megAI, KR1
; NAME/KEY: misc feature
; LOCATION: (17947)...(18207)
; OTHER INFORMATION: megAI, ACP1
; NAME/KEY: misc feature
; LOCATION: (18268)...(19548)
; OTHER INFORMATION: megAI, KS2
; NAME/KEY: misc feature
; LOCATION: (19876)...(20910)
; OTHER INFORMATION: megAI, AT2
; NAME/KEY: misc feature
; LOCATION: (22318)...(22575)
; OTHER INFORMATION: megAI, ACP2
; NAME/KEY: CDS
; LOCATION: (22867)...(33555)
; OTHER INFORMATION: megAI, SEQ ID NO: 14= translated amino acid sequence
; NAME/KEY: misc feature
; LOCATION: (22957)...(24237)
; OTHER INFORMATION: megAI, KS3
; NAME/KEY: misc feature
; LOCATION: (24544)...(25581)
; OTHER INFORMATION: megAI, AT3
; NAME/KEY: misc feature
; LOCATION: (26230)...(26733)
; OTHER INFORMATION: megAI, KR3 (inactive)
; NAME/KEY: misc feature
; LOCATION: (26998)...(27258)
; OTHER INFORMATION: megAI, ACP3
; NAME/KEY: misc feature
; LOCATION: (27393)...(28590)
; OTHER INFORMATION: megAI, KS4
; NAME/KEY: misc feature
; LOCATION: (28897)...(29931)
; OTHER INFORMATION: megAI, AT4
; NAME/KEY: misc feature
; LOCATION: (29953)...(30477)
; OTHER INFORMATION: megAI, DH4
; NAME/KEY: misc feature
; LOCATION: (31396)...(32244)
; OTHER INFORMATION: megAI, ER4
; NAME/KEY: misc feature
; LOCATION: (32257)...(32799)
; OTHER INFORMATION: megAI, KR4
; NAME/KEY: misc feature
; LOCATION: (33052)...(33312)
; OTHER INFORMATION: megAI, ACP4
; NAME/KEY: CDS
; LOCATION: (33666)...(43271)
; OTHER INFORMATION: megAI, SEQ ID NO: 15= translated amino acid sequence
; NAME/KEY: misc feature
; LOCATION: (33780)...(35027)
; OTHER INFORMATION: megAI, KS5
; NAME/KEY: misc feature
; LOCATION: (35365)...(36419)
; OTHER INFORMATION: megAI, AT5
; NAME/KEY: misc feature
; LOCATION: (37068)...(37604)
; OTHER INFORMATION: megAI, KR5
; NAME/KEY: misc feature
; LOCATION: (37860)...(38120)
; OTHER INFORMATION: megAI, ACP5
; NAME/KEY: misc feature
; LOCATION: (38187)...(39470)
; OTHER INFORMATION: megAI, KS6
; NAME/KEY: misc feature
; LOCATION: (39795)...(40811)

```

```

; OTHER INFORMATION: megAIII, AT6
; NAME/KEY: misc feature
; LOCATION: (41406)...(41936)
; OTHER INFORMATION: megAIII, KR6
; NAME/KEY: misc feature
; LOCATION: (42168)...(42425)
; OTHER INFORMATION: megAIII, ACP6
; NAME/KEY: misc feature
; LOCATION: (42585)...(43271)
; OTHER INFORMATION: megAIII, TE
; NAME/KEY: CDS
; LOCATION: (43268)...(44344)
; OTHER INFORMATION: megCII, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
; OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (44355)...(45623)
; OTHER INFORMATION: megCIII, desosaminyl transferase, desosamine glycosyltransferase
; OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (45620)...(46591)
; OTHER INFORMATION: megBII-2 (megBII), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
; OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (46660)...(47403)
; OTHER INFORMATION: megH, TEII; SEQ ID NO: 19= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (47411)...(47980)
; OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequ
; US-09-679-279-1

```

Query Match 4.2%; Score 21; DB 4; Length 47981;  
 Best Local Similarity 100.0%; Pred. No. 0.73;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 TTGACTCGCTGCGCGGTC 144  
 |||||  
 Db 22447 TTGACTCGCTGCGCGGTC 22467

RESULT 3  
 US-09-252-991A-10147  
 ; Sequence 10147, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 10147  
 ; LENGTH: 1410  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-10147

Query Match 4.0%; Score 20; DB 4; Length 1410;  
 Best Local Similarity 100.0%; Pred. No. 3.2;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 GGAGGCCGACGCGCCTGG 277  
 |||||  
 Db 459 GGAGGCCGACGCGCCTGG 478

RESULT 9  
 US-09-252-991A-9839/c

```
; Sequence 9839, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9839
; LENGTH: 1623
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-9839

Query Match 4.0%; Score 20; DB 4; Length 1623;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 GGAGCGCGACGACGGCTGG 277
Db 170 GGAGCGCGACGACGGCTGG 151

RESULT 10
US-09-252-991A-10214
; Sequence 10214, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10214
; LENGTH: 4026
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-10214

Query Match 4.0%; Score 20; DB 4; Length 4026;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 GGAGCGCGACGACGGCTGG 277
Db 3746 GGAGCGCGACGACGGCTGG 3765

RESULT 11
US-07-642-734C-3
; Sequence 3, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
```

```
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734C
; FILING DATE: 17-JAN-91
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dancikers, Andreas M
; REGISTRATION NUMBER: 32652
; REFERENCE/DOCKET NUMBER: 4952.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
; STRAIN: NRRL 238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..10722 /codon start= 19
; OTHER INFORMATION: /function= "gene eryA"
; OTHER INFORMATION: /product= "eryA ORE2 encoding modules 3 & 4 for
; OTHER INFORMATION: 6-deoxyerythronolide B"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 19..4470
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 97..1482
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1693..2670
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3406..3921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 4171..4428
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 4471..10722
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 4"
; FEATURE:
; NAME/KEY: misc feature
```



```

; LOCATION: 4471..5847
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 6054..7026
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 4"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 7165..9216
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: dehydratase and enoylreductase domains m"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 9433..9984
; OTHER INFORMATION: /function= "approximate span
; OTHER INFORMATION: beta-ketoreductase of module 4"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10225..10483
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 4"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10723..20235
; OTHER INFORMATION: /codon_start= 10723
; OTHER INFORMATION: /function= "gene =eryA"
; OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
; OTHER INFORMATION: 6-deoxyerythronolide B formatio"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10723..15165
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10831..12174
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 12379..13350
; OTHER INFORMATION: /function= "approximatr span of
; OTHER INFORMATION: acyltransferase domain of module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 14062..14610
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase of module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 14857..15114
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 15166..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 15172..16569
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 16768..17721
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 18379..18921

```

```

; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 19149..19398
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 19492..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
; US-07-642-734C-3

Query Match 4.0%; Score 20; DB 1; Length 20235;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 124 TTGACTCGCTGGCGCGGT 143
| | | | | | | | | | | | | | | | | | | |
Db 10354 TTGACTCGCTGGCGCGGT 10373

RESULT 12
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstock
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/POCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
; STRAIN: NREL 238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..10722
; OTHER INFORMATION: /codon_start= 19

```

OTHER INFORMATION: /function= "gene eryA"  
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for  
OTHER INFORMATION: 6-deoxyerythronolide B"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19..4470  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 97..1482  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1693..2670  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 3406..3921  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4171..4428  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4471..10722  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4471..5847  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 6054..7026  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 7165..9216  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: dehydratase and enoylreductase domains m"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 9433..9984  
OTHER INFORMATION: /function= "approximate span  
OTHER INFORMATION: beta-ketoreductase of module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10225..10483  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 4"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 10723..20235  
OTHER INFORMATION: /codon\_start= 10723  
OTHER INFORMATION: /function= "gene =eryA"  
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6  
OTHER INFORMATION: 6-deoxyerythronolide B formatio"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10723..15165  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10831..12174

OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 12379..13350  
OTHER INFORMATION: /function= "approximatr span of  
OTHER INFORMATION: acyltransferase domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 14062..14610  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 14857..15114  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 15166..20235  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 15172..16569  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 16768..17721  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 18379..18921  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19149..19398  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19492..20235  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: thioesterase domain of module 6"  
US-08-439-009A-3  
Query Match 4.0%; Score 20; DB 3; Length 20235;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 124 TTCGACTCGCTGCGCGCGGT 143  
Db 10354 TTCGACTCGCTGCGCGCGGT 10373  
RESULT 13  
US-08-804-227C-1  
Sequence 1, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kunstoss, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS

```
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..36067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
; US-08-804-227C-1

Query Match 4.0%; Score 20; DB 2; Length 43280;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 169 GCCACCGGGCTGGCGCTGCC 188
Db 13893 GCCACCGGGCTGGCGCTGCC 13912

RESULT 14
US-08-981-189B-16
; Sequence 16, Application US/08981189B
; Patent No. 6214797
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: UROCORTIN PEPTIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FITCH, EVEN, TABIN & FLANNERY
; STREET: 120 S. LaSalle Street, Suite 1600
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/981,189B
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Keowl, Wayne A.
; REGISTRATION NUMBER: 33,923
; REFERENCE/DOCKET NUMBER: 106.101.187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
;

; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..36067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
; US-08-804-227C-1

Query Match 3.8%; Score 19; DB 3; Length 375;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 CGCGGAGCAGACCGCATC 30
Db 333 CGCGGAGCAGACCGCATC 351

RESULT 15
US-08-866-340-24
; Sequence 24, Application US/08866340
; Patent No. 6020318
; GENERAL INFORMATION:
; APPLICANT: Szyf, Moshe
; APPLICANT: Bigey, Pascal
; APPLICANT: Ramchandani, Shyam
; TITLE OF INVENTION: DNA METHYLTRANSFERASE GENOMIC
; TITLE OF INVENTION: SEQUENCES AND ANTISENSE OLIGONUCLEOTIDES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States of America
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,340
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Keowl, Wayne A.
; REGISTRATION NUMBER: 33,923
; REFERENCE/DOCKET NUMBER: 106.101.187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
;
```

```

; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 451 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: linear
;   MOLECULE TYPE: other nucleic acid
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
; US-08-866-340-24

Query Match          3.8%; Score 19; DB 3; Length 451;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 CATCCTGCTGAAGCTGGTC 45
Db 200 CATCCTGCTGAAGCTGGTC 218

RESULT 16
US-09-103-875-30
; Sequence 30, Application US/09103875A
; Patent No. 6221849
; GENERAL INFORMATION:
; APPLICANT: Szyf, Moshe
; APPLICANT: Bigey, Pascal
; APPLICANT: Ramchandani, Shyam
; TITLE OF INVENTION: DNA METHYLTRANSFERASE GENOMIC SEQUENCES AND ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES
; FILE REFERENCE: 106101.194
; CURRENT APPLICATION NUMBER: US/09/103,875A
; CURRENT FILING DATE: 1998-06-24
; EARLIER APPLICATION NUMBER: 60/069,865
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 08/866,340
; EARLIER FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-103-875-30

Query Match          3.8%; Score 19; DB 3; Length 452;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 CATCCTGCTGAAGCTGGTC 45
Db 201 CATCCTGCTGAAGCTGGTC 219

RESULT 17
US-09-452-239-45
; Sequence 45, Application US/09452239
; Patent No. 6465229
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Fader, Gary M.
; APPLICANT: Cahoon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase
; FILE REFERENCE: BB1284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239
; CURRENT FILING DATE: 1998-12-01
; EARLIER APPLICATION NUMBER: 60/110,594
; EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 45
; LENGTH: 953
; TYPE: DNA

```

```

; ORGANISM: Triticum aestivum
; US-09-452-239-45

Query Match          3.8%; Score 19; DB 4; Length 953;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 CTGCTGAAGCTGGTCCGCG 49
Db 591 CTGCTGAAGCTGGTCCGCG 609

RESULT 18
US-09-452-239-43
; Sequence 43, Application US/09452239
; Patent No. 6465229
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Fader, Gary M.
; APPLICANT: Cahoon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase
; FILE REFERENCE: BB1284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239
; CURRENT FILING DATE: 1999-12-01
; EARLIER APPLICATION NUMBER: 60/110,594
; EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 43
; LENGTH: 1049
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (352)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (948)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (992)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (994)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (999)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1003)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1038)
; US-09-452-239-43

Query Match          3.8%; Score 19; DB 4; Length 1049;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 CTGCTGAAGCTGGTCCGCG 49
Db 591 CTGCTGAAGCTGGTCCGCG 609

RESULT 19
US-08-578-590-1
; Sequence 1, Application US/08578590
; Patent No. 5817499
; GENERAL INFORMATION:
; APPLICANT: Dalboge, Henrik
; APPLICANT: Christgau, Stephan
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kofod, Lene V.

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ;  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ;  
 ; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 8972  
 ; LENGTH: 717  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-8972

Query Match 3.6%; Score 18; DB 4; Length 717;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 CCGAAGCGGCCCGCGCG 427  
 DB 665 CCGAAGCGGCCCGCGCG 682

RESULT 24  
 US-09-252-991A-8861  
 ; Sequence 8861, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 8861  
 ; LENGTH: 819  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-8861

Query Match 3.6%; Score 18; DB 4; Length 819;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 CCGAAGCGGCCCGCGCG 427  
 DB 549 CCGAAGCGGCCCGCGCG 566

RESULT 25  
 US-09-252-991A-8752  
 ; Sequence 8752, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 8752  
 ; LENGTH: 1656  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-8752

Query Match 3.6%; Score 18; DB 4; Length 1656;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 CCGAAGCGGCCCGCGCG 427  
 DB 97 CCGAAGCGGCCCGCGCG 114

RESULT 26  
 US-08-399-646-3  
 ; Sequence 3, Application US/08399646  
 ; Patent No. 5556781  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KUBOTA, Michio  
 ; APPLICANT: TSUSAKI, Keiji  
 ; APPLICANT: HATTORI, Kazuo  
 ; APPLICANT: SUGIMOTO, Toshiyuki  
 ; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
 ; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
 ; NUMBER OF SEQUENCES: 18  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: BROWDY AND NEIMARK  
 ; STREET: 419 Seventh Street, N.W., Suite 300  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20004  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/399,646  
 ; FILING DATE: 07-MAR-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 59834  
 ; FILING DATE: 07-MAR-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 59840  
 ; FILING DATE: 07-MAR-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BROWDY, Roger L.  
 ; REGISTRATION NUMBER: 25,618  
 ; REFERENCE/DOCKET NUMBER: KUBOTA-5  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-628-5197  
 ; TELEFAX: 202-737-3528  
 ; TELEX: 248633  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1791 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cdna  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..1791  
 US-08-399-646-3

Query Match 3.6%; Score 18; DB 1; Length 1791;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTTGACTCGTGGCGCG 140  
 DB 1044 CTTGACTCGTGGCGCG 1061

RESULT 27  
US-08-607-321-3  
; Sequence 3, Application US/08607321  
; Patent No. 5716813  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/607,321  
; FILING DATE: 26-FEB-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,646  
; FILING DATE: 07-MAR-1995  
; APPLICATION NUMBER: JP 59834  
; FILING DATE: 07-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 59840  
; FILING DATE: 07-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: KUBOTA=5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1791 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1791  
US-08-607-321-3

Query Match 3.6%; Score 18; DB 1; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTTGACTCGTGGCGGC 140  
Db 1044 CTTGACTCGTGGCGGC 1061

RESULT 28  
US-08-961-240-3  
; Sequence 3, Application US/08961240  
; Patent No. 5830715  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji

; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,240  
; FILING DATE: 30-OCT-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,646  
; FILING DATE: 07-MAR-1995  
; APPLICATION NUMBER: JP 59834  
; FILING DATE: 07-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 59840  
; FILING DATE: 07-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: KUBOTA=5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1791 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1791  
US-08-961-240-3

Query Match 3.6%; Score 18; DB 2; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTTGACTCGTGGCGGC 140  
Db 1044 CTTGACTCGTGGCGGC 1061

RESULT 29  
US-08-605-501-3  
; Sequence 3, Application US/08605501  
; Patent No. 5834287  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300



CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605.501  
FILING DATE: 26-FEB-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/399,646  
FILING DATE: 07-MAR-1995  
APPLICATION NUMBER: JP 59834  
FILING DATE: 07-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 59840  
FILING DATE: 07-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: KUBOTA=5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1791 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1791  
US-08-605-501-3

Query Match 3.6%; Score 18; DB 2; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTTGACTCGTGGCGC 140  
|||||  
Db 1044 CTTGACTCGTGGCGC 1061

RESULT 30  
US-08-399-646-13  
Sequence 13, Application US/08399646  
Patent No. 5556781  
GENERAL INFORMATION:  
APPLICANT: KUBOTA, Michio  
APPLICANT: TSUSAKI, Keiji  
APPLICANT: HATTORI, Kazuko  
APPLICANT: SUGIMOTO, Toshiyuki  
TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/399,646  
FILING DATE: 07-MAR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 59834  
FILING DATE: 07-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 59840  
FILING DATE: 07-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: KUBOTA=5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2056 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 90..1883  
US-08-399-646-13

Query Match 3.6%; Score 18; DB 1; Length 2056;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTTGACTCGTGGCGC 140  
|||||  
Db 1136 CTTGACTCGTGGCGC 1153

RESULT 31  
US-08-607-321-13  
Sequence 13, Application US/08607321  
Patent No. 5716813  
GENERAL INFORMATION:  
APPLICANT: KUBOTA, Michio  
APPLICANT: TSUSAKI, Keiji  
APPLICANT: HATTORI, Kazuko  
APPLICANT: SUGIMOTO, Toshiyuki  
TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/607,321  
FILING DATE: 26-FEB-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/399,646  
FILING DATE: 07-MAR-1995  
APPLICATION NUMBER: JP 59834  
FILING DATE: 07-MAR-1994

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 59840
; FILING DATE: 07-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, ROGER L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: KUBOTA=5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2056 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..1883
; US-08-961-240-13
;
Query Match 3.6%; Score 18; DB 1; Length 2056;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 CTTGCACTCGTGGCGC 140
Db 1136 CTTGCACTCGTGGCGC 1153

RESULT 32
US-08-961-240-13
; Sequence 13, Application US/08961240
; Patent No. 5830715
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: HATTORI, Kazuko
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,240
; FILING DATE: 30-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,646
; FILING DATE: 07-MAR-1995
; APPLICATION NUMBER: JP 59834
; FILING DATE: 07-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 59840
; FILING DATE: 07-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, ROGER L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: KUBOTA=5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2056 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
;
Qy 123 CTTGCACTCGTGGCGC 140
Db 1136 CTTGCACTCGTGGCGC 1153

RESULT 33
US-08-605-501-13
; Sequence 13, Application US/08605501
; Patent No. 5834287
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: HATTORI, Kazuko
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605,501
; FILING DATE: 26-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,646
; FILING DATE: 07-MAR-1995
; APPLICATION NUMBER: JP 59834
; FILING DATE: 07-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 59840
; FILING DATE: 07-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, ROGER L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: KUBOTA=5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2056 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
;
Qy 123 CTTGCACTCGTGGCGC 140
Db 1136 CTTGCACTCGTGGCGC 1153
```

TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 90..1883  
US-08-605-501-13

Query Match 3.6%; Score 18; DB 2; Length 2056;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTTGCGCTGCTGGCGC 140  
Db 1136 CTTGCGCTGCTGGCGC 1153

## RESULT 34

US-09-252-991A-793  
Sequence 793, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 793  
LENGTH: 2508  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-793

Query Match 3.6%; Score 18; DB 4; Length 2508;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 GAGGCGCTGCTGGCTAC 237  
Db 1561 GAGGCGCTGCTGGCTAC 1578

## RESULT 35

US-09-252-991A-828/c  
Sequence 828, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 828  
LENGTH: 2577  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-828

Query Match 3.6%; Score 18; DB 4; Length 2577;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 GAGGCGCTGCTGGCTAC 237  
Db 1800 GAGGCGCTGCTGGCTAC 1783

## RESULT 36

US-08-764-233A-1  
Sequence 1, Application US/08764233A  
Patent No. 5716849  
GENERAL INFORMATION:  
APPLICANT: Ligon, James M.  
APPLICANT: Schupp, Thomas  
APPLICANT: Beck, James J.  
APPLICANT: Hill, Dwight S.  
APPLICANT: Neff, Snezana  
APPLICANT: Ryals, John A.  
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,233A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/729,214  
FILING DATE: 09-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: 1506/CIP6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 49377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Sorangium cellulosum  
IMMEDIATE SOURCE:  
CLONE: p98/1, pJL3, and pVKM15  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 383..760  
OTHER INFORMATION: /product= "SorR"  
OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKSs tha  
OTHER INFORMATION: the reductase domains of type I PKSs such as eryA from  
OTHER INFORMATION: Saccharopolyspora erythraea."  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 927..19874  
OTHER INFORMATION: /product= "SorA"  
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSs tha  
OTHER INFORMATION: are known to be involved in the synthesis of polyketide  
OTHER INFORMATION: compounds."  
FEATURE:



```

; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-09-154-083-13

Query Match          3.4%; Score 17; DB 3; Length 561;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 TCGGCTTCGACTCGCTG 135
    |||
Db 295 TCGGCTTCGACTCGCTG 279
    |||

RESULT 43
US-09-146-221-9/c
; Sequence 9, Application US/09146221
; Patent No. 6180895
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; APPLICANT: Wildung, Mark R
; APPLICANT: McCaskill, David G
; TITLE OF INVENTION: Nucleic and Amino Acid Sequences Relating to a No. 6190895el
; TITLE OF INVENTION: Transketolase, and Methods for the Expression Thereof
; FILE REFERENCE: No. 6190895el transketolase from peppermint
; CURRENT APPLICATION NUMBER: US/09/146,221
; CURRENT FILING DATE: 1998-09-01
; EARLIER APPLICATION NUMBER: 60/056,033
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Mentha piperita
US-09-146-221-9

Query Match          3.4%; Score 17; DB 3; Length 617;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 CGGCGCCGAGGCATCG 94
    |||
Db 89 CGGCGCCGAGGCATCG 73
    |||

RESULT 44
US-09-252-991A-5281
; Sequence 5281, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5281
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5281

```

```

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5395
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5395

Query Match          3.4%; Score 17; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CGAGCAGAACCGCATCC 31
    |||
Db 347 CGAGCAGAACCGCATCC 363
    |||

RESULT 41
US-09-252-991A-12682
; Sequence 12682, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12682
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12682

Query Match          3.4%; Score 17; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 CCTCGCCGACACCGGCA 382
    |||
Db 127 CCTCGCCGACACCGGCA 143
    |||

RESULT 42
US-09-154-083-13/c
; Sequence 13, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz

```

```
Query Match      3.4%; Score 17; DB 4; Length 723;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      215 CCCCGGAGGCGCTGGTC 231
      |||
Db      513 CCCCGGAGGCGCTGGTC 529

RESULT 45
US-09-252-991A-5345
; Sequence 5345, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5345
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5345

Query Match      3.4%; Score 17; DB 4; Length 744;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      215 CCCCGGAGGCGCTGGTC 231
      |||
Db      544 CCCCGGAGGCGCTGGTC 560

Search completed: June 16, 2004, 19:32:22
Job time : 71.3333 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:56:30 ; Search time 241 Seconds  
(without alignments)

9488.873 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543

Perfect score: 501

Sequence: 1 cgggtcccgacgcgagca.....ccctcgggacagagctga 501

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2998549 seqs, 2282253817 residues

Word size : 15

Total number of hits satisfying chosen parameters: 2932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:\*\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq\*
- 3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	24	4.8	10056	15	US-10-156-761-411
2	24	4.8	86941	17	US-10-461-194-2
3	24	4.8	100000	15	US-10-156-761-15103
C 4	24	4.8	9025608	15	US-10-156-761-1
C 5	23	4.6	1552	13	US-10-425-114-1096
C 6	23	4.6	1583	13	US-10-425-114-962
7	23	4.6	14643	15	US-10-156-761-931
8	23	4.6	31422	17	US-10-204-862A-2
9	22	4.4	84428	17	US-10-229-148B-1
10	21	4.2	10692	15	US-10-156-761-414
C 11	20	4.0	1089	16	US-10-369-493-31550
12	20	4.0	2640	16	US-10-369-493-44289
13	20	4.0	4551	13	US-10-282-122A-30490
14	20	4.0	14520	15	US-10-156-761-2885

c	15	20	4.0	16596	15	US-10-156-761-930	Sequence 930, App
c	16	20	4.0	31422	17	US-10-204-862A-2	Sequence 2, Appli
c	17	20	4.0	125746	15	US-10-156-761-15102	Sequence 15102, A
c	18	20	4.0	9025608	15	US-10-156-761-1	Sequence 1, Appli
c	19	19	3.8	375	10	US-09-818-009-16	Sequence 16, Appl
c	20	19	3.8	485	15	US-10-198-846-141	Sequence 141, App
c	21	19	3.8	726	15	US-10-156-761-5550	Sequence 5550, Ap
c	22	19	3.8	953	9	US-09-452-239-45	Sequence 45, Appl
c	23	19	3.8	996	15	US-10-156-761-5938	Sequence 5938, Ap
c	24	19	3.8	1044	15	US-10-156-761-7258	Sequence 7258, Ap
c	25	19	3.8	1049	9	US-09-452-239-43	Sequence 43, Appl
c	26	19	3.8	1293	15	US-10-156-761-5634	Sequence 5634, Ap
c	27	19	3.8	1893	9	US-09-815-242-8019	Sequence 8019, Ap
c	28	19	3.8	1893	13	US-10-282-122A-30672	Sequence 30672, A
c	29	19	3.8	2000	15	US-10-198-846-10290	Sequence 10290, A
c	30	19	3.8	2040	17	US-10-250-613-22	Sequence 22, Appl
c	31	19	3.8	2856	16	US-10-369-493-37369	Sequence 37369, A
c	32	19	3.8	3081	17	US-10-450-826-14	Sequence 14, Appl
c	33	19	3.8	5434	9	US-09-962-436-285	Sequence 285, App
c	34	19	3.8	5434	9	US-09-954-531-172	Sequence 172, App
c	35	19	3.8	5434	9	US-09-954-531-381	Sequence 381, App
c	36	19	3.8	5434	15	US-10-144-577-1	Sequence 1, Appli
c	37	19	3.8	11058	15	US-10-156-761-3629	Sequence 3629, Ap
c	38	18	3.6	520	13	US-10-425-114-23847	Sequence 23847, A
c	39	18	3.6	546	13	US-10-389-647-154	Sequence 154, App
c	40	18	3.6	855	16	US-10-369-493-42814	Sequence 42814, A
c	41	18	3.6	960	13	US-10-282-122A-36804	Sequence 36804, A
c	42	18	3.6	963	9	US-09-815-242-9687	Sequence 9687, Ap
c	43	18	3.6	963	13	US-10-282-122A-38849	Sequence 38849, A
c	44	18	3.6	963	13	US-10-282-122A-39841	Sequence 39841, A
c	45	18	3.6	1245	15	US-10-156-761-2391	Sequence 2391, Ap

#### ALIGNMENTS

##### RESULT 1

US-10-156-761-411  
; Sequence 411, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 411  
; LENGTH: 10056  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(10056)  
US-10-156-761-411

Query Match 4.8%; Score 24; DB 15; Length 10056;  
Best Local Similarity 100.0%; Pred. No. 0.027; 0; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 0;

QY 184 CTGCCGCGACGCTGATCTCGAC 207  
Db 4456 CTGCCGCGACGCTGATCTCGAC 4479



APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```

; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700101596_FLI
US-10-425-114-1096

```

```
Query Match      4.6%; Score 23; DB 13; Length 1552;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Query Match 4.6%; Score 23; DB 15; Length 14643;  
Best Local Similarity 100.0%; Pred. No. 0.079;  
Matches 23; Conservative 0; Mismatches 0; Indels 0

QY 298 CGGCGAGTCCTCGCGGCCGTGCC 320  
|||  
Db 294 CGGCGAGTCCTCGCGGCCGTGCC 272

Qy 193 ACGCTGATCTTCGACTACCCAC 215  
Db 4645 ACGCTGATCTTCGACTACCCAC 4667

RESULT 6  
US-10-425-114-962/c  
; Sequence 962, Application US/10425114  
; Publication No. US20040034888A1

RESULT 8  
US-10-204-862A-2  
; Sequence 2, Application US/10204862A  
; Publication No. US20040101936A1

```

: GENERAL INFORMATION:
: APPLICANT: Liu, Jingtong
: APPLICANT: Zhou, Yihua
: APPLICANT: Kovalic, David K.
: APPLICANT: Screen, Steven E
: APPLICANT: Tabaska, Jack E
: APPLICANT: Cao, Yongwei
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21(53313)B
: CURRENT APPLICATION NUMBER: US/10/425,114
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 73128

```

/ APPLICANT: ENDO, HIROFUMI  
 / APPLICANT: YAMAGUCHI, HIROYUKI  
 / APPLICANT: KANDA, YUTAKA  
 / APPLICANT: HASHIMOTO, SHINICHI  
 / APPLICANT: OMURA, SATOSHI  
 / APPLICANT: IKEDA, HARUO  
 / TITLE OF INVENTION: PROCESS FOR PRODUCING AVERMECTIN DERIVATIVE  
 / FILE REFERENCE: 468-32  
 / CURRENT APPLICATION NUMBER: US/10/204, 862A  
 / CURRENT FILING DATE: 2003-12-26  
 / PRIOR APPLICATION NUMBER: JP 00/047405  
 / PRIOR FILING DATE: 2000-02-24  
 / NUMBER OF SEQ ID NOS: 11  
 / SOFTWARE: PatentIn Ver. 2.0

; LENGTH: 1583  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 ; FEATURE:

```

; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 2
; LENGTH: 31422
; TYPE: DNA
;

```

FEATURE: 700097375\_FLI  
OTHER INFORMATION: Clone ID: 700097375\_FLI  
US-10-425-114-962

```

; ORGANISM: Streptomyces ave
;
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: (1)..(14643)
;
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: (14824)..(31419)
US-10-204-862A-2

```

Query Match 4.6%; Score 23; DB 13; Length 1583;  
Best Local Similarity 100.0%; Pred.No. 0.13;  
Matches 23; Conservative 0; Mismatches 0; Indels

LOCATION: (1)..(14643)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (14824)..(31419)  
 S-10-204-862A-2

Query Match 4.6%; Score 23; DB 17; Length 31422;  
 Best Local Similarity 100.0%; Pred. NO. 0.067;  
 Matched 23; Conservative 0; Indels 0

Qy 298 CGGCGAGTCCTCGCGGCCGTGCC 320  
|||  
Db 320 CGGCGAGTCCTCGCGGCCGTGCC 298

US-10-204-862A-2	Query Match	4.6%;	Score 23
	Best Local Similarity	100.0%;	Prod. N
	Matches	23; Conservative	0; Mism
QY	193	ACGCTGATCTTCGACTACCCAC	215
Dh	4645	ACGCTGATCTTCGACTACCCAC	4667

RESULT 7  
US-10-156-761-931  
; Sequence 931, Application US/10156761  
; Publication No. US20030119018A1

	RESULT 7	
QY	US-10-156-761-931	
	; Sequence 931, Application US/10156761	
D8	; Publication No. US20030119018A1	
	; GENERAL INFORMATION:	
	; APPLICANT: OMURA, SATOSHI	
	; APPLICANT: IKEDA, HARUO	
	; APPLICANT: ISHIKAWA, JUN	
	; APPLICANT: HORIKAWA, HIROSHI	
	; APPLICANT: SHIBA, TADAYOSHI	
	; APPLICANT: SAKAKI, YOSHIYUKI	
	; APPLICANT: HATTORI, MASAHIRA	
	; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES	
	; FILE REFERENCE: 249-262	
	; CURRENT APPLICATION NUMBER: US/10/156,761	
	; CURRENT FILING DATE: 2002-05-29	
	; PRIOR APPLICATION NUMBER: JP 2001-204089	
	; PRIOR FILING DATE: 2001-05-30	
	; PRIOR APPLICATION NUMBER: JP 2001-272697	
	; PRIOR FILING DATE: 2001-08-02	

  

	RESULT 9	
QY	US-10-229-148B-1	
	; Sequence 1, Application US/10229148B	
D8	; Publication No. US20040091975A1	
	; GENERAL INFORMATION:	
	; APPLICANT: Meiji Seika Kaisha, Ltd.	
	; TITLE OF INVENTION: Midecamycin biosynthetic genes	
	; FILE REFERENCE: 138451 US	
	; CURRENT APPLICATION NUMBER: US/10/229,148B	
	; CURRENT FILING DATE: 2002-08-28	
	; PRIOR APPLICATION NUMBER: 210516/2002	
	; PRIOR FILING DATE: 2002-07-19	
	; NUMBER OF SEQ ID NOS: 48	
	; SOFTWARE: Patent In Ver. 2.1	

  

QY	193	ACGCTGATCTTCGACTACCCAC	215
D8	4645	ACGCTGATCTTCGACTACCCAC	4667

```

; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 931
; LENGTH: 14643
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
;
; SEQ ID NO 1
; LENGTH: 8428
; TYPE: DNA
; ORGANISM: Streptomyces mycarofaciens
; FEATURE:

```

```
; SEQ ID NO 1
; LENGTH: 84428
; TYPE: DNA
; ORGANISM: Str
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: Complement((1)..(675))
; OTHER INFORMATION: ORF42 (fragment)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((1168)..(2202))
; OTHER INFORMATION: ORF41
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((2220)..(3215))
; OTHER INFORMATION: ORF40
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((3237)..(4691))
; OTHER INFORMATION: ORF39
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((4695)..(5948))
; OTHER INFORMATION: ORF38
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((6048)..(6629))
; OTHER INFORMATION: ORF37
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((6653)..(7945))
; OTHER INFORMATION: ORF36
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8149)..(9015)
; OTHER INFORMATION: ORF35
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9012)..(9335)
; OTHER INFORMATION: ORF34
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9328)..(10458)
; OTHER INFORMATION: ORF33
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10521)..(11603)
; OTHER INFORMATION: ORF32
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11729)..(12961)
; OTHER INFORMATION: ORF31
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((13016)..(14044))
; OTHER INFORMATION: ORF30
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((14074)..(15096))
; OTHER INFORMATION: ORF29
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15643)..(17466)
; OTHER INFORMATION: ORF28
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((17522)..(18895))
; OTHER INFORMATION: ORF27
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19063)..(20229)
; OTHER INFORMATION: ORF26
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((20307)..(21743))
; OTHER INFORMATION: ORF25
; NAME/KEY: CDS

; LOCATION: Complement((21733)..(22527))
; OTHER INFORMATION: ORF24
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((22534)..(23571))
; OTHER INFORMATION: ORF23
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((23555)..(24463))
; OTHER INFORMATION: ORF22
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((24460)..(25650))
; OTHER INFORMATION: ORF21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((25647)..(26105))
; OTHER INFORMATION: ORF20
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((26180)..(27391))
; OTHER INFORMATION: ORF19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27937)..(28983)
; OTHER INFORMATION: ORF18
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29244)..(42779)
; OTHER INFORMATION: ORF1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42823)..(48657)
; OTHER INFORMATION: ORF2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (48712)..(59802)
; OTHER INFORMATION: ORF3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (59850)..(64556)
; OTHER INFORMATION: ORF4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64687)..(70365)
; OTHER INFORMATION: ORF5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70365)..(71078)
; OTHER INFORMATION: ORF6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71113)..(72360)
; OTHER INFORMATION: ORF7
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72400)..(73665)
; OTHER INFORMATION: ORF8
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73694)..(75043)
; OTHER INFORMATION: ORF9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((75899)..(76570))
; OTHER INFORMATION: ORF10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((76602)..(77765))
; OTHER INFORMATION: ORF11
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78039)..(79313)
```

```
/ OTHER INFORMATION: ORF12
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: Complement((79391)..(81052))
/ OTHER INFORMATION: ORF13
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (81541)..(82356)
/ OTHER INFORMATION: ORF14
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (82760)..(83362)
/ OTHER INFORMATION: ORF15
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: Complement((83495)..(84142))
/ OTHER INFORMATION: ORF16
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (84329)..(84428)
/ OTHER INFORMATION: ORF17 (fragment)
US-10-229-148B-1

Query Match      4.4%; Score 22; DB 17; Length 84428;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 TCGGCTTCGACTCGCTGGCGC 140
      |||||
Db 64243 TCGGCTTCGACTCGCTGGCGC 64264

RESULT 10
US-10-156-761-414
/ Sequence 414, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ PRIOR FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 414
/ LENGTH: 10692
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(10692)
US-10-156-761-414

Query Match      4.2%; Score 21; DB 15; Length 10692;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 CTCGGCTGCCCGACGCTG 198
      |||||
Db 10357 CTCGGCTGCCCGACGCTG 10377

RESULT 11
US-10-369-493-31550/c

/ Sequence 31550, Application US/10369493
/ Publication No. US20030233675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ PRIOR FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 31550
/ LENGTH: 1089
/ TYPE: DNA
/ ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31550

Query Match      4.0%; Score 20; DB 16; Length 1089;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGACCGCATCC 31
      |||||
Db 208 CGCCGAGCAGACCGCATCC 189

RESULT 12
US-10-369-493-44289
/ Sequence 44289, Application US/10369493
/ Publication No. US20030233675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ CURRENT FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 44289
/ LENGTH: 2640
/ TYPE: DNA
/ ORGANISM: Rhodospseudomonas palustris
US-10-369-493-44289

Query Match      4.0%; Score 20; DB 16; Length 2640;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 CGCGCGCGAAGCATCGGCC 97
      |||||
Db 1683 CGCGCGCGAAGCATCGGCC 1702

RESULT 13
US-10-282-122A-30490
/ Sequence 30490, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
```

Thu Jun 17 11:20:19 2004

```

; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30490
; LENGTH: 4551
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30490

Query Match 4.0%; Score 20; DB 13; Length 4551;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 GGAGCGCGACGACGCGCTGG 277
Db 3600 GGAGCGCGACGACGCGCTGG 3619

RESULT 14
US-10-156-761-2885
; Sequence 2885, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2885
; LENGTH: 14520
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(16596)
US-10-156-761-2885

Query Match 4.0%; Score 20; DB 15; Length 14520;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 ACCGGGCTCGGCTGCCGC 191
Db 4270 ACCGGGCTCGGCTGCCGC 4289

RESULT 15
US-10-156-761-930/c
; Sequence 930, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 930
; LENGTH: 16596
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(16596)
US-10-156-761-930

Query Match 4.0%; Score 20; DB 15; Length 16596;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 CCTCGCGACACCGCACCG 385
Db 12387 CCTCGCGACACCGCACCG 12368

RESULT 16
US-10-204-862A-2/c
; Sequence 2, Application US/10204862A
; Publication No. US20040101936A1
; GENERAL INFORMATION:
; APPLICANT: ENDO, HIROFUMI
; APPLICANT: YAMAGUCHI, HIROYUKI
; APPLICANT: KANDA, YUTAKA
; APPLICANT: HASHIMOTO, SHINICHI
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; TITLE OF INVENTION: PROCESS FOR PRODUCING AVERMECTIN DERIVATIVE
; FILE REFERENCE: 468-32
; CURRENT APPLICATION NUMBER: US/10/204,862A
; CURRENT FILING DATE: 2003-12-26
; PRIOR APPLICATION NUMBER: JP 00/047405
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 11
```

```

; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(14520)
US-10-156-761-2885

Query Match 4.0%; Score 20; DB 15; Length 14520;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 ACCGGGCTCGGCTGCCGC 191
Db 4270 ACCGGGCTCGGCTGCCGC 4289

RESULT 15
US-10-156-761-930/c
; Sequence 930, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 930
; LENGTH: 16596
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(16596)
US-10-156-761-930

Query Match 4.0%; Score 20; DB 15; Length 16596;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 CCTCGCGACACCGCACCG 385
Db 12387 CCTCGCGACACCGCACCG 12368

RESULT 16
US-10-204-862A-2/c
; Sequence 2, Application US/10204862A
; Publication No. US20040101936A1
; GENERAL INFORMATION:
; APPLICANT: ENDO, HIROFUMI
; APPLICANT: YAMAGUCHI, HIROYUKI
; APPLICANT: KANDA, YUTAKA
; APPLICANT: HASHIMOTO, SHINICHI
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; TITLE OF INVENTION: PROCESS FOR PRODUCING AVERMECTIN DERIVATIVE
; FILE REFERENCE: 468-32
; CURRENT APPLICATION NUMBER: US/10/204,862A
; CURRENT FILING DATE: 2003-12-26
; PRIOR APPLICATION NUMBER: JP 00/047405
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 11
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 31422
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(14643)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (14824)..(31419)
US-10-204-862A-2

Query Match          4.0%; Score 20; DB 17; Length 31422;
Best Local Similarity 100.0%; Pred. No. 2.1; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 366 CCTGCGCGACACCGGCACCG 385
Db 27210 CCTGCGCGACACCGGCACCG 27191

RESULT 17
US-10-156-761-15102
; Sequence 15102, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15102
; LENGTH: 125746
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15102

Query Match          4.0%; Score 20; DB 15; Length 125746;
Best Local Similarity 100.0%; Pred. No. 1.6; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 ACCGGGCTGCGGCTGCCCGC 191
Db 48046 ACCGGGCTGCGGCTGCCCGC 48065

RESULT 18
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15102
; LENGTH: 125746
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15102

Query Match          4.0%; Score 20; DB 15; Length 125746;
Best Local Similarity 100.0%; Pred. No. 1.6; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 ACCGGGCTGCGGCTGCCCGC 191
Db 48046 ACCGGGCTGCGGCTGCCCGC 48065

RESULT 19
US-09-818-009-16
; Sequence 16, Application US/09818009
; Publication No. US20030032587A1
; GENERAL INFORMATION:
; APPLICANT: THE SALK INSTITUTE FOR BIOLOGICAL STUDIES
; TITLE OF INVENTION: UROCORTIN PEPTIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FITCH, EVEN, TABIN & FLANNERY
; STREET: 120 S. LaSalle Street, Suite 1600
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/818,009
; APPLICATION NUMBER: US/09/818,009
; FILING DATE: 26-Mar-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/981,189
; FILING DATE: 10-DEC-1997
; APPLICATION NUMBER: US 60/028,144
; FILING DATE: 13-JUN-1995
; APPLICATION NUMBER: US 60/002,223
; FILING DATE: 11-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Schumann, James J.
; REGISTRATION NUMBER: 20,856
; REFERENCE/DOCKET NUMBER: 57611
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-552-1311
; TELEFAX: 858-552-0095
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 375 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human Genomic Placental Library
```

```
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..375
; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-818-009-16
    Query Match          3.8%; Score 19; DB 10; Length 375;
    Best Local Similarity 100.0%; Pred. No. 17;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGAACCGCATC 30
Db 333 CGCCGAGCAGAACCGCATC 351

RESULT 20
US-10-198-846-141/c
; Sequence 141, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 141
; LENGTH: 485
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 380..402
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-141
    Query Match          3.8%; Score 19; DB 15; Length 485;
    Best Local Similarity 100.0%; Pred. No. 16;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGAACCGCATC 30
Db 181 CGCCGAGCAGAACCGCATC 163

RESULT 21
US-10-156-761-5550/c
; Sequence 550, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
```

```
;
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5550
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-156-761-5550
    Query Match          3.8%; Score 19; DB 15; Length 726;
    Best Local Similarity 100.0%; Pred. No. 14;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 GAAGACGACCTCCGGCGAG 304
Db 528 GAAGACGACCTCCGGCGAG 510

RESULT 22
US-09-452-239-45
; Sequence 45, Application US/09452239
; Patent No. US20020081693A1
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Fader, Gary M.
; APPLICANT: Cahoon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase
; FILE REFERENCE: BB1284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239
; CURRENT FILING DATE: 1999-12-01
; EARLIER APPLICATION NUMBER: 60/110,594
; EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 45
; LENGTH: 953
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-452-239-45
    Query Match          3.8%; Score 19; DB 9; Length 953;
    Best Local Similarity 100.0%; Pred. No. 14;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 CTGCTGAAGCTGGTCCGCG 49
Db 591 CTGCTGAAGCTGGTCCGCG 609

RESULT 23
US-10-156-761-5938
; Sequence 5938, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5938
; LENGTH: 996
```



```
;
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(996)
US-10-156-761-5938

Query Match          3.8%; Score 19; DB 15; Length 996;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 CGCGCCACCGGCTGCGG 183
Db 30 CGCGCCACCGGCTGCGG 48

RESULT 24
US-10-156-761-7258/c
; Sequence 7258, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7258
; LENGTH: 1044
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1044)
US-10-156-761-7258

Query Match          3.8%; Score 19; DB 15; Length 1044;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 GCGTGAAGTCTGCGGGA 260
Db 809 GCGTGAAGTCTGCGGGA 791

RESULT 25
US-09-452-239-43
; Sequence 43, Application US/09452239
; Patent No. US20020081693A1
; GENERAL INFORMATION:
; APPLICANT: Rafaleki, Antoni J.
; APPLICANT: Fader, Gary M.
; APPLICANT: Cahoon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase
; FILE REFERENCE: BB1284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239
; CURRENT FILING DATE: 1999-12-01
; EARLIER APPLICATION NUMBER: 60/110,594
; EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 43
; LENGTH: 1049

; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (352)
; NAME/KEY: unsure
; LOCATION: (948)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (992)
; NAME/KEY: unsure
; LOCATION: (994)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (999)
; NAME/KEY: unsure
; LOCATION: (1003)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1038)
US-09-452-239-43

Query Match          3.8%; Score 19; DB 9; Length 1049;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 CTGCTGAAGTGTGTCGCG 49
Db 591 CTGCTGAAGTGTGTCGCG 609

RESULT 26
US-10-156-761-5634
; Sequence 5634, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5634
; LENGTH: 1293
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1293)
US-10-156-761-5634

Query Match          3.8%; Score 19; DB 15; Length 1293;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 GCTGGACACGCTGCTCGGC 366
Db 99 GCTGGACACGCTGCTCGGC 117
```

```
RESULT 27
US-09-815-242-8019
; Sequence 8019, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8019
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1893)
US-09-815-242-8019
Query Match 3.8%; Score 19; DB 9; Length 1893;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 CGAAGGCATCGCCCGCGC 102
Db 807 CGAAGGCATCGCCCGCGC 825

RESULT 28
US-10-282-122A-30672
; Sequence 30672, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20

Qy 84 CGAAGGCATCGCCCGCGC 102
Db 807 CGAAGGCATCGCCCGCGC 825

US-10-198-846-10290
; Sequence 10290, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10290
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 1829, 1830, 1831, 1997, 1998, 1999, 2000
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10290
Query Match 3.8%; Score 19; DB 15; Length 2000;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 CGCCGAGCAGAACCGCATC 30
Db 807 CGCCGAGCAGAACCGCATC 30
```

```
Db 701 CGCCGAGCAGAACCGCATC 719
; LENGTH: 2856
; TYPE: DNA
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-37369

Query Match 3.8%; Score 19; DB 16; Length 2856;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 GCGCGGAGGCGATCGGCC 97
Db 2035 GCGCGGAGGCGATCGGCC 2053

RESULT 32
US-10-450-826-14
; Sequence 14, Application US/10450826
; Publication No. US2004010181A1
; GENERAL INFORMATION:
; APPLICANT: Ji, Darren
; APPLICANT: Axelrod, Douglas W.
; APPLICANT: Cook, Jonathon S.
; APPLICANT: Jaiswal, Neelam
; APPLICANT: Eistein, Richard
; APPLICANT: Houghton, Adam
; APPLICANT: Mertz, Lawrence
; TITLE OF INVENTION: Gene Expression Profiles Associated with Osteoblast Differentiation
; FILE REFERENCE: 044921-5039-WO
; CURRENT APPLICATION NUMBER: US/10/450,826
; CURRENT FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 60/255,882
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 60/285,691
; PRIOR FILING DATE: 2001-04-24
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 3081
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. AF038633
US-10-450-826-14

Query Match 3.8%; Score 19; DB 17; Length 3081;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGAACCGCATC 30
Db 2915 CGCCGAGCAGAACCGCATC 2933

RESULT 33
US-09-962-436-285
; Sequence 285, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 285
; LENGTH: 5434
; TYPE: DNA
```

```
Db 701 CGCCGAGCAGAACCGCATC 719
; LENGTH: 2856
; TYPE: DNA
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-37369

Query Match 3.8%; Score 19; DB 16; Length 2856;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 GCGCGGAGGCGATCGGCC 97
Db 2035 GCGCGGAGGCGATCGGCC 2053

RESULT 32
US-10-450-826-14
; Sequence 14, Application US/10450826
; Publication No. US2004010181A1
; GENERAL INFORMATION:
; APPLICANT: Ji, Darren
; APPLICANT: Axelrod, Douglas W.
; APPLICANT: Cook, Jonathon S.
; APPLICANT: Jaiswal, Neelam
; APPLICANT: Eistein, Richard
; APPLICANT: Houghton, Adam
; APPLICANT: Mertz, Lawrence
; TITLE OF INVENTION: Gene Expression Profiles Associated with Osteoblast Differentiation
; FILE REFERENCE: 044921-5039-WO
; CURRENT APPLICATION NUMBER: US/10/450,826
; CURRENT FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 60/255,882
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 60/285,691
; PRIOR FILING DATE: 2001-04-24
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 3081
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. AF038633
US-10-450-826-14

Query Match 3.8%; Score 19; DB 17; Length 3081;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGAACCGCATC 30
Db 2915 CGCCGAGCAGAACCGCATC 2933

RESULT 33
US-09-962-436-285
; Sequence 285, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 285
; LENGTH: 5434
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-09-962-436-285

Query Match          3.8%; Score 19; DB 9; Length 5434;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 CATCCTGCTGAAGCTGGTC 45
Db 3813 CATCCTGCTGAAGCTGGTC 3831

RESULT 34
US-09-954-531-172
; Sequence 172, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; TITLE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 172
; LENGTH: 5434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-172

Query Match          3.8%; Score 19; DB 9; Length 5434;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 CATCCTGCTGAAGCTGGTC 45
Db 3813 CATCCTGCTGAAGCTGGTC 3831

RESULT 35
US-09-954-531-381
; Sequence 381, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; TITLE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
```

```
; SEQ ID NO 381
; LENGTH: 5434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-381

Query Match          3.8%; Score 19; DB 9; Length 5434;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 CATCCTGCTGAAGCTGGTC 45
Db 3813 CATCCTGCTGAAGCTGGTC 3831

RESULT 36
US-10-144-577-1
; Sequence 1, Application US/10144577
; Publication No. US20030083292A1
; GENERAL INFORMATION:
; APPLICANT: MacLeod, Alan Robert
; TITLE OF INVENTION: Inhibitors of DNA Methyltransferase Isoforms
; FILE REFERENCE: MET-005
; CURRENT APPLICATION NUMBER: US/10/144,577
; CURRENT FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: US 60/290,202
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,212
; PRIOR FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-144-577-1

Query Match          3.8%; Score 19; DB 15; Length 5434;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 CATCCTGCTGAAGCTGGTC 45
Db 3813 CATCCTGCTGAAGCTGGTC 3831

RESULT 37
US-10-156-761-3629
; Sequence 3629, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3629
; LENGTH: 11058
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
```

```
; LOCATION: (1)..(11058)
US-10-156-761-3629

Query Match          3.8%; Score 19; DB 15; Length 11058;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 GCTCGGCTCGCGACACC 378
Db 10500 GCTCGGCTCGCGACACC 10518

RESULT 38
US-10-425-114-23847
; Sequence 23847, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 23847
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Chlorella sorokiniana
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3602-017-G8_FLI
US-10-425-114-23847

Query Match          3.6%; Score 18; DB 13; Length 520;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 GAGCGGGCGTGTGGAC 354
Db 30 GAGCGGGCGTGTGGAC 47

RESULT 39
US-10-389-647-154
; Sequence 154, Application US/10389647
; Publication No. US20040033549A1
; GENERAL INFORMATION:
; APPLICANT: GREENBERG, E. Peter
; APPLICANT: SCHUSTER, Martin
; APPLICANT: LOSTROH, Candl
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
; FILE REFERENCE: UIZ-038CP
; CURRENT APPLICATION NUMBER: US/10/389,647
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 09/653730
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/153022
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 710
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 154
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-389-647-154

Query Match          3.6%; Score 18; DB 13; Length 546;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 CCGAAGCGCGCCCGCGCG 427
Db 494 CCGAAGCGCGCCCGCGCG 511

RESULT 40
US-10-369-493-42814
; Sequence 42814, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 42814
; LENGTH: 855
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-10-369-493-42814

Query Match          3.6%; Score 18; DB 16; Length 855;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 339 GCGCGGGCGTGTGGACAC 356
Db 432 GCGCGGGCGTGTGGACAC 449

RESULT 41
US-10-282-122A-36804
; Sequence 36804, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
```

; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 36804  
; LENGTH: 960  
; TYPE: DNA  
; ORGANISM: Salmonella paratyphi A  
US-10-282-122A-36804

Query Match 3.6%; Score 18; DB 13; Length 960;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 CGACGGTGTCTCGGCACA 76  
|||||  
Db 734 CGACGGTGTCTCGGCACA 751

RESULT 42  
US-09-815-242-9687  
; Sequence 9687, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9687  
; LENGTH: 963  
; TYPE: DNA  
; ORGANISM: Salmonella typhi  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(963)  
US-09-815-242-9687  
Query Match 3.6%; Score 18; DB 9; Length 963;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 CGACGGTGTCTCGGCACA 76

Db 734 CGACGGTGTCTCGGCACA 751  
|||||  
RESULT 43  
US-10-282-122A-38849  
; Sequence 38849, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 38849  
; LENGTH: 963  
; TYPE: DNA  
; ORGANISM: Salmonella typhimurium  
US-10-282-122A-38849

Query Match 3.6%; Score 18; DB 13; Length 963;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 CGACGGTGTCTCGGCACA 76  
|||||  
Db 734 CGACGGTGTCTCGGCACA 751

RESULT 44  
US-10-282-122A-39841  
; Sequence 39841, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 39841  
; LENGTH: 963  
; TYPE: DNA  
; ORGANISM: Salmonella typhi  
US-10-282-122A-39841

Query Match 3.6%; Score 18; DB 13; Length 963;  
Best Local Similarity 100.0%; Pred.No. 43;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 59 CGACGGTGCTCGGCACA 76  
|||||  
Db 734 CGACGGTGCTCGGCACA 751

RESULT 45  
US-10-156-761-2391  
; Sequence 2391, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2391  
; LENGTH: 1245  
; TYPE: DNA

; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1245)  
US-10-156-761-2391  
Query Match 3.6%; Score 18; DB 15; Length 1245;  
Best Local Similarity 100.0%; Pred.No. 41;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 356 CGCTGCTCGGCTCGCCG 373  
|||||  
Db 374 CGCTGCTCGGCTCGCCG 391  
Search completed: June 16, 2004, 20:02:40  
Job time : 1336 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:37:20 ; Search time 2102.33 Seconds  
(without alignments)  
7116.356 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543

Perfect score: 501

Sequence: 1 gcggtcccgagcgagca.....cctcggcgagcgagtgta 501

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 15

Total number of hits satisfying chosen parameters: 16108

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hcc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_nam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gsal:\*

29: gb\_gsal2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23	4.6	479	14	CF650732
C 2	23	4.6	501	12	B1273373
C 3	23	4.6	601	9	AI973519
C 4	23	4.6	742	29	CG384197

C 5	23	4.6	820	29	CG297728
C 6	23	4.6	829	28	BZ731737
C 7	23	4.6	857	28	CG381148
C 8	23	4.6	858	28	BZ731730
C 9	23	4.6	872	29	CG092377
C 10	23	4.6	876	29	CG289214
C 11	23	4.6	900	28	CC381145
C 12	23	4.6	918	29	CG384209
C 13	23	4.6	974	29	CG289224
C 14	23	4.6	1889	11	AY103974
C 15	22	4.4	340	14	DI5897
C 16	21	4.2	629	14	CA741791
C 17	21	4.2	1189	10	AW982464
C 18	20	4.0	354	14	CF770975
C 19	20	4.0	375	14	CF759576
C 20	20	4.0	491	12	BM371243
C 21	20	4.0	683	12	BI957736
C 22	20	4.0	702	28	CC346318
C 23	20	4.0	717	12	BG962173
C 24	20	4.0	770	29	CG670138
C 25	20	4.0	783	29	CG382731
C 26	20	4.0	802	29	CG382740
C 27	20	4.0	849	28	BZ649718
C 28	20	4.0	1035	10	BE791387
C 29	19	3.8	112	14	CB927303
C 30	19	3.8	211	10	BF822655
C 31	19	3.8	219	9	AI352333
C 32	19	3.8	219	9	AI631150
C 33	19	3.8	220	10	BF773138
C 34	19	3.8	251	9	AI971801
C 35	19	3.8	346	9	AI716696
C 36	19	3.8	349	10	BE415413
C 37	19	3.8	356	10	BF941825
C 38	19	3.8	358	14	CA646451
C 39	19	3.8	379	14	H24268
C 40	19	3.8	384	10	BF523813
C 41	19	3.8	399	14	CA614966
C 42	19	3.8	410	10	BE415414
C 43	19	3.8	414	10	BF474477
C 44	19	3.8	439	10	AW294748
C 45	19	3.8	453	12	BM358573

ALIGNMENTS

RESULT 1  
CF650732/c  
LOCUS  
DEFINITION  
3530\_1\_92\_1\_E03.x.1\_3530 - Full length cDNA library created by  
Invitrogen from multiple tissues Zea mays cDNA, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Zea mays  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
1 (bases 1 to 479)  
Maize ESTs from various cDNA libraries sequenced at Stanford  
University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 3530\_1\_92\_1 row: E column: 03.  
Location/Qualifiers

CF650732 479 bp mRNA linear EST 02-OCT-2003  
3530\_1\_92\_1\_E03.x.1\_3530 - Full length cDNA library created by  
Invitrogen from multiple tissues Zea mays cDNA, mRNA sequence.

CF650732.1 GI:37425965  
EST.  
Zea mays

```

source
1. .479
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/tissue_type="multiple"
/dev_stage="varies by tissue"
/lab_host="DH10B"
/clone_lib="3530 - Full length cDNA library created by
Invitrogen from multiple tissues"
/notes="Organ: silks, husks, ears, pollen, shoot tips,
leaf, root tips, whole seed, embryo; Vector: pCMV-SPORT
6.1; Site 1: EcoRV; Site 2: NotI; Maize Gene Discovery
Project contracted with Invitrogen to produce a
normalized, full length library in a pSPORT vector. This
is a Gateway compatible vector, permitting clone movement
to new vector backbones for expression in diverse host
cells using recombination rather than restriction enzymes.
Details of the vector and sequencing primers are available
at ZmDB in the EST library description tables. poly(A)+
mRNA was prepared by Invitrogen, and equimolar amounts of
RNA from each of the 12 tissue samples were mixed together
for selection of mRNA with a 5' cap. After synthesis of
cDNA, a normalization step was conducted against the
mixture of RNA sources. This step effected a 20X to 80X
reduction in common transcript types. Tissues prepared: 1.
just emerging silks; 2. inner husks from ears of sample
#1; 3. 20 dap aleurone; 4. immature tassels, stages from
1-2 mm to 1-2 cm; 5. 2 mm to 2 cm ears; 6. pollen; 7. 1 cm
vegetative shoot tips from 15 day old seedlings; all
leaves with an expanded or partially expanded sheath
were removed; 8. mature leaf tissue; 9. 0.5 cm long root
tips from 15 day old seedlings; 10. 10 dap whole seed; 11.
12 dap endosperm and embryo; 12. 17 dap endosperm and
embryo. All of the sequenced clones in project 3530 will
be archived at the University of Arizona along with the
unigene clones from the Maize Gene Discovery EST
sequencing projects. Clones can be ordered through the
ZmDB web site or directly from the University of Arizona
(http://www.genome.arizona.edu/orders/). High density
filters containing over 18,000 clones can also be ordered
from the University of Arizona."

ORIGIN
Query Match 4.6%; Score 23; DB 14; Length 479;
Best Local Similarity 100.0%; Pred.No. 46;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGGCGAGTCCTCGCGCGCGTGCC 320
|||||
Db 354 CGGCGAGTCCTCGCGCGCGTGCC 332

RESULT 2
Bi273373/c
LOCUS 949024B09.y1 949 - Juvenile leaf and shoot cDNA from Steve Moose
DEFINITION Zea mays cDNA, mRNA sequence.
ACCESSION Bi273373
VERSION Bi273373.1 GI:14883487
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 501)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 496027 row: F column: 10.
Location/Qualifiers
1. .601
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"

source
1. .501
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W64A"
/db_xref="taxon:4577"
/tissue_type="immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/notes="Organ: juvenile vegetative shoots; Vector:
PAD-GAL4-2.1; Site 1: EcoRI; Site 2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcoRI (5') and XhoI (3') directional cloning into lambda
Hybridap vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imbibing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing - includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf
primordia, 9 days after sowing. 4. Partially expanded and
greening leaves 4-5 at 13 days after sowing."

ORIGIN
Query Match 4.6%; Score 23; DB 12; Length 501;
Best Local Similarity 100.0%; Pred.No. 46;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGGCGAGTCCTCGCGCGCGTGCC 320
|||||
Db 319 CGGCGAGTCCTCGCGCGCGTGCC 297

RESULT 3
AI973519/c
LOCUS 496027F10.x1 496 - stressed shoot cDNA library from Wang/Bohnert
DEFINITION lab Zea mays cDNA, mRNA sequence.
ACCESSION AI973519
VERSION AI973519.1 GI:5770345
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 601)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 496027 row: F column: 10.
Location/Qualifiers
1. .601
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"

```

```

Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 496024 row: B column: 09.
Location/Qualifiers
1. .501
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W64A"
/db_xref="taxon:4577"
/tissue_type="immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/notes="Organ: juvenile vegetative shoots; Vector:
PAD-GAL4-2.1; Site 1: EcoRI; Site 2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcoRI (5') and XhoI (3') directional cloning into lambda
Hybridap vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imbibing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing - includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf
primordia, 9 days after sowing. 4. Partially expanded and
greening leaves 4-5 at 13 days after sowing."

ORIGIN
Query Match 4.6%; Score 23; DB 12; Length 501;
Best Local Similarity 100.0%; Pred.No. 46;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGGCGAGTCCTCGCGCGCGTGCC 320
|||||
Db 319 CGGCGAGTCCTCGCGCGCGTGCC 297

RESULT 3
AI973519/c
LOCUS 496027F10.x1 496 - stressed shoot cDNA library from Wang/Bohnert
DEFINITION lab Zea mays cDNA, mRNA sequence.
ACCESSION AI973519
VERSION AI973519.1 GI:5770345
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 601)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 496027 row: F column: 10.
Location/Qualifiers
1. .601
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"

```

```

/db_xref="taxon:4577"
/tissue_type="seedling"
/dev_stages="salt stress"
/lab_host="E.coli XL Gold"
/clone_lib="496 - stressed shoot cDNA library from
Wang/Bohnert lab"
/note="Organ: shoot; Vector: pBluescriptII SK(+) XR;
Wang/Bohnert"

ORIGIN
Query Match 4.6%; Score 23; DB 9; Length 601;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCTCGCGCGCGTGC 320
Db 288 CGCGAGTCTCGCGCGCGTGC 266

RESULT 4
CG384197
LOCUS CG384197 742 bp DNA linear GSS 26-AUG-2003
DEFINITION OG08R89TH ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0681P10,
genomic survey sequence.
ACCESSION CG384197
VERSION CG384197.1 GI:34301464
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 742)
Whiteaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: OG08R89TV
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1..742
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBMA0681P10"
/clone_lib="ZM 0.7 1.5 KB"
/note="Vector: pBCKS; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 4.6%; Score 23; DB 29; Length 820;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCTCGCGCGCGTGC 320
Db 506 CGCGAGTCTCGCGCGCGTGC 484

RESULT 6
BZ731737/c
LOCUS BZ731737 829 bp DNA linear GSS 03-MAR-2003
DEFINITION OGFAS22TM ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0241P09,
genomic survey sequence.
ACCESSION BZ731737
VERSION BZ731737.1 GI:28705867
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 829)
Whiteaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: OGFAS29TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
Location/Qualifiers

ORIGIN
Query Match 4.6%; Score 23; DB 29; Length 742;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCTCGCGCGCGTGC 320
Db 147 CGCGAGTCTCGCGCGCGTGC 169

RESULT 5
CG297728/c
LOCUS CG297728 820 bp DNA linear GSS 25-AUG-2003
DEFINITION OG3BA37TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0766H01,
genomic survey sequence.

```

```

source
1. .829
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBMA0241F09"
/clone_lib="ZM 0.7_1.5_KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 4.6%; Score 23; DB 28; Length 829;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGGCGAGTCCTCGCGCGCGTCC 320
Db 415 CGGCGAGTCCTCGCGCGCGTCC 393

RESULT 7
CC381148 857 bp DNA linear GSS 19-MAY-2003
LOCUS PHNW75TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTA494M05,
DEFINITION genomic survey sequence.
ACCESSION CC381148
VERSION CC381148.1 GI:30858013
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 857)
AUTHORS Whitelaw C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PFHW75TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. .857
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBTA494M05"
/clone_lib="ZM 0.6_1.0_KB"
/notes="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

ORIGIN
Query Match 4.6%; Score 23; DB 28; Length 857;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGGCGAGTCCTCGCGCGCGTCC 320
Db 495 CGGCGAGTCCTCGCGCGCGTCC 517

RESULT 8
BZ731730 858 bp DNA linear GSS 03-MAR-2003
LOCUS BZ731730
DEFINITION OGFAS29TC ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMA0241F09,

```

```

genomic survey sequence.
BZ731730
VERSION BZ731730.1 GI:28705853
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 858)
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nummer, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGFAS29TM
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. .858
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBMA0241F09"
/clone_lib="ZM 0.7_1.5_KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 4.6%; Score 23; DB 28; Length 858;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGGCGAGTCCTCGCGCGCGTCC 320
Db 777 CGGCGAGTCCTCGCGCGCGTCC 799

RESULT 9
CG092377 872 bp DNA linear GSS 20-AUG-2003
LOCUS PUFWT10TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTA0750B20,
DEFINITION genomic survey sequence.
ACCESSION CG092377
VERSION CG092377.1 GI:33974671
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 872)
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUFWT10TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.

```

```

FEATURES
  source      Location/Qualifiers
    1..872
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBTA0750B20"
      /notes="Vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high
      Cot selected genomic DNA library"

ORIGIN
Query Match      4.6%; Score 23; DB 29; Length 872;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCTCGCGCGGTGCC 320
Db 278 CGCGAGTCTCGCGCGGTGCC 300

RESULT 10
CG289214/c
LOCUS
DEFINITION
  OGWMK56TH ZM 0.7 1.5 KB DNA linear GSS 25-AUG-2003
  genomic survey sequence.
ACCESSION
  CG289214
VERSION
  CG289214.1 GI:34203428
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 876)
  Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
  Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
  Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Other_GSSs: OGWMK56TV
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TR
  Class: sheared ends.
  Location/Qualifiers
    1..876
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBMA0617J15"
      /notes="Vector: PBCSk-; Site 1: HincII; 0.7-1.5 kb
      methylation filtered genomic DNA library"

ORIGIN
Query Match      4.6%; Score 23; DB 29; Length 876;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCTCGCGCGGTGCC 320
Db 413 CGCGAGTCTCGCGCGGTGCC 391

RESULT 11
CC381145/c
LOCUS
DEFINITION
  PUHNW75TB ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTA494M05,
  genomic survey sequence.
ACCESSION
  CC381145
VERSION
  CC381145.1 GI:30858008
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 900)
  Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
  Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
  Bennetzen,J.
  Maize Genomics Consortium
  Unpublished (2003)
  Other_GSSs: PUHNW75TD
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TR
  Class: sheared ends.
  Location/Qualifiers
    1..900
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBTA494M05"
      /clone_lib="ZM 0.6 1.0 KB"
      /notes="Vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high
      Cot selected genomic DNA library"

ORIGIN
Query Match      4.6%; Score 23; DB 28; Length 900;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCTCGCGCGGTGCC 320
Db 408 CGCGAGTCTCGCGCGGTGCC 386

RESULT 12
CG384209/c
LOCUS
DEFINITION
  CG0BR89TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0661P10,
  genomic survey sequence.
ACCESSION
  CG384209
VERSION
  CG384209.1 GI:34301476
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 918)
  Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
  Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
  Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Other_GSSs: CG0BR89TH
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TF

```

```

Class: sheared ends.
Location/Qualifiers
1. .918
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBMA061P10"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 4.6%; Score 23; DB 29; Length 918;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCTCTCGCGCGTGC 320
Db 878 CGCGAGTCTCTCGCGCGTGC 856

RESULT 13
CG289224
LOCUS
DEFINITION OGMW56TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0617J15,
genomic survey sequence.
ACCESSION CG289224
VERSION CG289224.1 GI:34203438
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 974)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nurnberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
OTHER_GSSs: OGMW56TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. .974
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBMA0617J15"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 4.6%; Score 23; DB 29; Length 974;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCTCTCGCGCGTGC 320
Db 942 CGCGAGTCTCTCGCGCGTGC 964

RESULT 14
AY103974/c

```

PUBMED 7987417  
 COMMENT Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@abr.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/  
 PROJECT = "RGP"

Seq primer: oligo(dT).  
 Location/Qualifiers  
 1..340  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="C1483A"  
 /clone\_lib="Rice callus"  
 /note="Vector: pBluescript II SK+; Site 1: SalI; Site 2:  
 NotI; cDNA prepared from rice callus mRNAs by using  
 oligo(dT) as a primer and ligating to the SalI-NotI site  
 of pBluescript II SK+ phagemid."

ORIGIN  
 Query Match 4.4%; Score 22; DB 14; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 CGGTCTCGGCACAGCGCGC 83  
 DB 187 CGGTCTCGGCACAGCGCGC 208

RESULT 16  
 LOCUS CA741791 629 bp mRNA linear EST 26-NOV-2002  
 DEFINITION wialc.pk003.f19 wialc Triticum aestivum cDNA clone wialc.pk003.f19  
 5' end, mRNA sequence.  
 ACCESSION CA741791  
 VERSION CA741791.1 GI:25557614  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooideae; Triticeae; Triticum.  
 1 (bases 1 to 629)  
 Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,  
 Miao, G., Caraher, N. and Hanafey, M.K.  
 DuPont Wheat cDNA Sequence  
 Unpublished (2002)  
 Contact: Scott V. Tingey  
 Crop Genetics  
 E. I. DuPont de Nemours and Company  
 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
 Tel: 302-631-2602  
 Fax: 302-631-2607  
 Email: Scott.V.Tingey@USA.dupont.com  
 Seq primer: M13.  
 Location/Qualifiers  
 1..629  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4565"  
 /clone="wialc.pk003.f19"  
 /tissue\_type="anthers"  
 /lab\_host="DH10B"  
 /clone\_lib="wialc"  
 /note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:  
 XhoI; Wheat (Triticum aestivum, Hi Line) immature anthers"

ORIGIN  
 Query Match 4.2%; Score 21; DB 14; Length 629;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 GCCTCGCGACACCGGACCG 385  
 DB 134 GCCTCGCGACACCGGACCG 154

RESULT 17  
 LOCUS AW982464/c 1189 bp mRNA linear EST 22-OCT-2001  
 DEFINITION HVSMEG0003F11f Hordeum vulgare pre-anthesis spike EST library  
 HVCNA0008 (white to yellow anther) Hordeum vulgare subsp. vulgare  
 cDNA clone HVSMEG0003F11f, mRNA sequence.  
 ACCESSION AW982464  
 VERSION AW982464.3 GI:16316777  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare subsp. vulgare  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooideae; Triticeae; Hordeum.  
 1 (bases 1 to 1189)  
 Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D.,  
 Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W.,  
 Fenton, R.D., Close, S.J., Oates, R. and Main, D.  
 Development of a genetically and physically anchored EST resource  
 for barley genomics: Morex pre-anthesis spike cDNA library  
 Unpublished (2001)  
 On Jun 2, 2000 this sequence version replaced gi:13153149.  
 Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Total hg bases = 181  
 Seq primer: AATTACCTCTACTAAGG  
 High quality sequence stop: 834.  
 Location/Qualifiers  
 1..1189  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Morex"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="HVSMEG0003F11f"  
 /tissue\_type="pre-anthesis spike"  
 /lab\_host="SOLR"  
 /clone\_lib="Hordeum vulgare pre-anthesis spike EST library  
 HVCNA0008 (white to yellow anther)"  
 /note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
 Plants were grown in the greenhouse at the University of  
 California, Riverside (Fenton, SJ Close, TJ Close). Whole  
 spike with awns trimmed were collected at white, green and  
 yellow anther stages (Fenton). Total RNA was prepared from  
 each pool, equal quantities of all three RNA pools were  
 combined, poly(A) RNA was purified from the mixture, one  
 primary unamplified cDNA library was made, and 1 million  
 pfu were in vivo excised to give pBluescript SK(-) cDNA  
 phagemids. These steps were performed in the TJ Close lab  
 (Choi) at the University of California, Riverside.  
 Phagemids were plated and picked at the Clemson University  
 Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins  
 and Wing) plasmid DNA preparations, DNA sequencing and  
 sequence analysis were performed at CUGI (Wing, Yu,  
 Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence  
 has been trimmed to remove vector sequence and contains a  
 minimum of 100 bases of phred value 20 or above. For more  
 details on library preparation and sequence analysis see  
 http://www.genome.clemson.edu/projects/barley. To order  
 this clone see http://www.genome.clemson.edu/orders Also

FEATURES  
 source



see Close TJ, Wing R, Kleinjohs A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html) "

## ORIGIN

Query Match 4.2%; Score 21; DB 10; Length 1189;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 CTGGCCACAGCGCGCGGAA 87  
|||||  
Db 933 CTGGCCACAGCGCGCGGAA 913

## RESULT 18

CF770975 354 bp mRNA linear EST 20-OCT-2003  
DEFINITION DSBF1\_12\_C09.g1\_A010 Drought-stressed before flowering Sorghum  
bicolor cDNA clone DSBF1\_12\_C09\_A010 3', mRNA sequence.

ACCESSION CF770975.1 GI:37755892

VERSION EST.

KEYWORDS Sorghum bicolor (sorghum)

SOURCE Sorghum bicolor

ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 354)

AUTHORS Cordonnier-Pratt, M.-M., Zhang, D., McCartor, K., Nguyen, H.T. and  
Pratt, L.H.

TITLE An EST Database from Sorghum: Subtracted pre-flowering drought  
stressed leaf tissues

JOURNAL Unpublished (2003)

COMMENT Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed at Texas Tech University by Deshui Zhang in the  
Laboratory of Dr. Henry Nguyen. Sequencing was done in the  
Laboratory for Genomics and Bioinformatics, University of Georgia.  
Sequence ends have been trimmed to exclude vector and regions below  
phred quality 16. Three-prime sequences are presented as their  
reverse complement and have been trimmed to exclude polyA.

Seq primer: T7 (TAATACGACTCACTATAGG)

POLYA=Yes.

## FEATURES

## source

1..354 Location/Qualifiers

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/cultivar="Tx7000"

/db\_xref="taxon:4558"

/clone="DSBF1\_12\_C09\_A010"

/dev\_stage="Pre-flowering"

/lab\_host="ElectroMax DH10B (BRL)"

/clone\_lib="brought-stressed before flowering"

/notes="Organ: Leaf; Vector: pBluescriptSK-; Site: 1: XhoI;

Site 2: EcoRI; The library was prepared from polyA+ RNA

from leaves harvested from pre-flowering, drought-stressed

Sorghum bicolor, cv. TX7000. Double-stranded cDNA was

cloned unidirectionally using the Unizap system from

Stratagene. After amplification, the library was

subtracted by re-association hybridization. Inserts can be

excised with XhoI and EcoRI."

## ORIGIN

Query Match 4.0%; Score 20; DB 14; Length 354;  
Best Local Similarity 100.0%; Pred. No. 7.9e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 TCAGCGCCGTCGCGTCGCC 327  
|||||  
Db 19 TCAGCGCCGTCGCGTCGCC 38

## RESULT 19

CF759576

LOCUS

DEFINITION

DSAF1\_46\_H11.b1\_A011 Drought-stressed after flowering Sorghum

bicolor cDNA clone DSAF1\_46\_H11\_A011 5', mRNA sequence.

ACCESSION CF759576

VERSION EST.

KEYWORDS Sorghum bicolor (sorghum)

SOURCE Sorghum bicolor

ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD

clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 375)

AUTHORS Cordonnier-Pratt, M.-M., Zhang, D., McCartor, K., Nguyen, H.T. and

Pratt, L.H.

TITLE An EST Database from Sorghum: Subtracted post-flowering drought

stressed leaf tissues

JOURNAL Unpublished (2003)

COMMENT Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed at Texas Tech University by Deshui Zhang and

Jianhang Jia in the laboratory of Dr. Henry Nguyen. Sequencing was

done in the Laboratory for Genomics and Bioinformatics, University

of Georgia. Sequence ends have been trimmed to exclude vector and

regions below phred quality 16. Three-prime sequences are presented

as their reverse complement and have been trimmed to exclude polyA.

Seq primer: JENREV (CAGGAACAGCTATGACC).

Location/Qualifiers

1..375

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/cultivar="B35"

/db\_xref="taxon:4558"

/clone="DSAF1\_46\_H11\_A011"

/dev\_stage="Post-flowering"

/lab\_host="ElectroMax DH10B (BRL)"

/clone\_lib="Drought-stressed after flowering"

/notes="Organ: Leaf; Vector: pBluescriptSK-; Site: 1: XhoI;

Site 2: EcoRI; The library was prepared from polyA+ RNA

from leaves harvested from post-flowering,

drought-stressed Sorghum bicolor, cv. B35. Double-stranded

cDNA was cloned unidirectionally using the Unizap system

from Stratagene. After amplification, the library was

subtracted by re-association hybridization. Inserts can be

excised with XhoI and EcoRI."

ORIGIN

Query Match 4.0%; Score 20; DB 14; Length 375;

Best Local Similarity 100.0%; Pred. No. 7.9e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 TCAGCGCCGTCGCGTCGCC 327

|||||

Db 19 TCAGCGCCGTCGCGTCGCC 38

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

ACCESSION BM371243  
 VERSION BM371243.2 GI:21949498  
 SOURCE EST.  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Hordeum vulgare subsp. vulgare  
 Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Hordeum.  
 1 (bases 1 to 491)  
 REFERENCE Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,  
 AUTHORS Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.  
 Development of Barley Transcriptome Resources  
 Unpublished (2001)  
 TITLE On Jan 10, 2002 this sequence version replaced gi:18114633.  
 JOURNAL Contact: Waugh R, Marshall DF  
 COMMENT Genome Dynamics/Computational Biology  
 Scottish Crop Research Institute  
 Invergowrie, Dundee, DD2 5DA, Scotland, UK  
 Tel: 00 44 1382 562731  
 Fax: 00 44 1382 562426  
 Email: est@scri.sari.ac.uk  
 All sequence has a Phred quality score of 20 or over  
 Seq primer: M13 reverse  
 FEATURES  
 source  
 Location/Qualifiers  
 1..491  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Optic"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="EBro04\_SQ003\_P12"  
 /tissue\_type="root"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="root, 3 week, salt-stressed, cv Optic, EBro04"  
 /note="Vector: pSPOR11; Site 1: Sal I; Site 2: Not I;  
 Non-normalised library, directionally cloned into pSPOR11.  
 Derived from roots of 3 week old salt stressed barley  
 plants. Developed as part of the barley transcriptome  
 resources of BSRC/SERAD funded cereal IGF (Investigating  
 Gene Function) project."  
 ORIGIN  
 Query Match 4.0%; Score 20; DB 12; Length 491;  
 Best Local Similarity 100.0%; Pred. No. 8.1e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 416 CGGCCCCGGCGCGCGACGAC 435  
 DB 356 CGGCCCCGGCGCGCGACGAC 337  
 RESULT 21  
 B1957736/c 683 bp mRNA linear EST 22-OCT-2001  
 LOCUS HVSME0010P11f Hordeum vulgare rachis EST library HVcDNA0015  
 DEFINITION (normal) Hordeum vulgare subsp. vulgare cDNA clone HVSME0010P11f,  
 mRNA sequence.  
 B1957736  
 B1957736.1 GI:16308991  
 SOURCE EST.  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Hordeum vulgare subsp. vulgare  
 Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Hordeum.  
 1 (bases 1 to 683)  
 REFERENCE Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Chin, A., Begum, D.,  
 AUTHORS Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T.,  
 Simmons, J., Oates, R. and Main, D.  
 Development of a genetically and physically anchored EST resource  
 for barley genomics: Morex rachis cDNA library  
 Unpublished (2001)  
 TITLE  
 JOURNAL

## COMMENT

Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Total hg bases = 186  
 Seq primer: AATTACCTCTACTAAGG  
 High quality sequence stop: 289.

## FEATURES

source  
 Location/Qualifiers  
 1..683  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Morex"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="HVSME0010P11f"  
 /tissue\_type="Rachis"  
 /lab\_host="TJC121"  
 /clone\_lib="Hordeum vulgare rachis EST library HVcDNA0015  
 (normal)"  
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; Plants were grown at Washington State University,  
 Pullman, WA in a greenhouse, the rachises were excised and  
 frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close  
 lab at the University of California, Riverside total RNA  
 was prepared, poly(A) was purified, one primary  
 unamplified cDNA library was made, and 1 million pfu were  
 in vivo excised to give pBluescript SK(-) cDNA phagemids  
 (Chin). Phagemids were plated and picked at the Clemson  
 University Genomics Institute (CUGI) (Begum, Palmer,  
 Frisch, Atkins and Wing). Plasmid DNA preparations, DNA  
 sequencing and sequence analysis were performed at CUGI  
 (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The  
 sequence has been trimmed to remove vector sequence and  
 contains a minimum of 100 bases of phred value 20 or  
 above. For more details on library preparation and  
 sequence analysis see  
 http://www.genome.clemson.edu/projects/barley. To order  
 this clone see http://www.genome.clemson.edu/orders Also  
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
 Genetically and physically anchored EST resources for  
 barley genomics. Barley Genetics Newsletter 31:29-30.  
 (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

## ORIGIN

Query Match 4.0%; Score 20; DB 12; Length 683;  
 Best Local Similarity 100.0%; Pred. No. 8.4e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 258 GGAGGCCGACGCGCGCTGG 277  
 DB 221 GGAGGCCGACGCGCGCTGG 202  
 RESULT 22  
 CC346318  
 LOCUS OG0BD84TH ZM 0.7 1.5 KB Zea mays genomic clone ZMvBma0343M24,  
 DEFINITION genomic survey sequence.  
 ACCESSION CC346318  
 VERSION CC346318.1 GI:30815725  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 702)  
 REFERENCE Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
 AUTHORS Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE Consortium for Maize Genomics  
JOURNAL Unpublished (2002)  
COMMENT Contact: Cathy Whitelaw  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TR  
Class: sheared ends.

Location/Qualifiers

1. .702  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0343M24"  
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 4.0%; Score 20; DB 28; Length 702;  
Best Local Similarity 100.0%; Pred. No. 8.4e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 359 TGCTCGGCTCGCGACACC 378  
|||||  
Db 138 TGCTCGGCTCGCGACACC 157  
|||||

## RESULT 23

BG962173/c 717 bp mRNA linear EST 12-JUN-2001  
LOCUS 602826940F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4981881 5',  
mRNA sequence.

ACCESSION BG962173  
VERSION BG962173.1 GI:14349810  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 717)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM10983 row: p column: 10  
High quality sequence stop: 249.

## FEATURES

source  
1. .717  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4981881"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP Co24"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site.1: NotI;  
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.6 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

Query Match 4.0%; Score 20; DB 12; Length 717;  
Best Local Similarity 100.0%; Pred. No. 8.4e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 CGCGGCCACCGGCTCGGC 184  
|||||  
Db 98 CGCGGCCACCGGCTCGGC 79  
|||||

## RESULT 24

CC670138 770 bp DNA linear GSS 19-JUN-2003  
LOCUS OGLCF27TV ZM.0.7.1.5 KB Zea mays genomic clone ZMMBMA0321E06,  
genomic survey sequence.

ACCESSION CC670138  
VERSION CC670138.1 GI:32074380  
KEYWORDS GSS.

## SOURCE

ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

AUTHORS 1 (bases 1 to 770)  
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

## TITLE

Consortium for Maize Genomics

## JOURNAL

## COMMENT

Unpublished (2002)

Other GSSs: OGLCF27TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

## FEATURES

source  
1. .770  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0321E06"  
/clone\_lib="ZM.0.7.1.5 KB"  
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 4.0%; Score 20; DB 29; Length 770;  
Best Local Similarity 100.0%; Pred. No. 8.5e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 359 TGCTCGGCTCGCGACACC 378  
|||||  
Db 254 TGCTCGGCTCGCGACACC 273  
|||||

## RESULT 25

CG382731/c 783 bp DNA linear GSS 26-AUG-2003  
LOCUS OG1BK25TH ZM.0.7.1.5 KB Zea mays genomic clone ZMMBMA0724E01,  
genomic survey sequence.

ACCESSION CG382731  
VERSION CG382731.1 GI:34299998  
KEYWORDS GSS.

## SOURCE

ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 783)  
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,

Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

# TITLE JOURNAL COMMENT

Unpublished (2002)  
Other\_GSSs: OG1BK25TV  
Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: sheared ends.

## FEATURES source

1..783  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0724E01"  
/clone\_lib="ZM\_0.7\_1.5\_KB"  
/note="Vector: pBCSK-; Site\_1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 4.0%; Score 20; DB 29; Length 783;  
Best Local Similarity 100.0%; Pred.No. 8.5e+02; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

QY 359 TGCTCGGCTCGCGACACC 378  
|||||  
Db 431 TGCTCGGCTCGCGACACC 412  
|||||

RESULT 26  
CG382740  
LOCUS  
DEFINITION  
CG382740 802 bp DNA linear GSS 26-AUG-2003  
genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Zea mays  
Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE AUTHORS

1 (bases 1 to 802)  
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

## TITLE JOURNAL COMMENT

Unpublished (2002)  
Other\_GSSs: OG1BK25TH  
Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: sheared ends.

## FEATURES source

1..802  
/organism="Zea mays"  
/mol\_type="Genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0724E01"  
/clone\_lib="ZM\_0.7\_1.5\_KB"  
/note="Vector: pBCSK-; Site\_1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 4.0%; Score 20; DB 29; Length 802;  
Best Local Similarity 100.0%; Pred.No. 8.5e+02; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

QY 359 TGCTCGGCTCGCGACACC 378  
|||||  
Db 713 TGCTCGGCTCGCGACACC 732  
|||||

## RESULT 27 BZ649718/c

LOCUS  
DEFINITION  
BZ649718 849 bp DNA linear GSS 29-JAN-2003  
genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Zea mays  
Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE AUTHORS

1 (bases 1 to 849)  
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

## TITLE JOURNAL COMMENT

Unpublished (2002)  
Other\_GSSs: OGCBQ11TM  
Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: sheared ends.

## FEATURES source

1..849  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0135B21"  
/clone\_lib="ZM\_0.7\_1.5\_KB"  
/note="Vector: pBCSK-; Site\_1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 4.0%; Score 20; DB 28; Length 849;  
Best Local Similarity 100.0%; Pred.No. 8.5e+02; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

QY 410 CCGAAGCGGCGCGCGCC 429  
|||||  
Db 740 CCGAAGCGGCGCGCGCC 721  
|||||

## RESULT 28 BE791387

LOCUS  
DEFINITION  
BE791387 1035 bp mRNA linear EST 20-SEP-2000  
601583888F1 NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:3938540 5',  
mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE AUTHORS

1 (bases 1 to 1035)  
NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov

Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: L1CM795 row: 0 column: 21  
High quality sequence stop: 444.

## FEATURES

source

Location/Qualifiers  
1. .1035  
/organism="Homo sapiens"  
/mol type="mRNA"  
/db xref="taxon:9606"  
/clone="IMAGE:3938540"  
/tissue type="small cell carcinoma"  
/cell line="MGC3"  
/lab host="DH10B (phage-resistant)"  
/clone lib="NIH\_MGC.7"  
/note="Organ: lung; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match 4.0%; Score 20; DB 10; Length 1035;  
Best Local Similarity 100.0%; Pred. No. 8.7e+02; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 276 GGACGGCGGAGACGACC 295

Db 776 GGACGGCGGAGACGACC 795

## RESULT 29

CB927303/c

LOCUS

DEFINITION ABAL\_14\_B04.g1 A012 Abscisic acid-treated seedlings Sorghum bicolor  
cDNA clone ABAL\_14\_B04\_A012 5', mRNA sequence.

ACCESSION

CB927303.1

VERSION

GI:30163574

KEYWORDS

EST.

SOURCE

ORGANISM

Sorghum bicolor (sorghum)

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD

Clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 112)

Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Buchanan,C.D., Eastman,A. and Pratt,L.H.

An EST database from Sorghum: ABAL-treated seedlings

Unpublished (2003)

Other ESTs: ABAL\_14\_B04.b1\_A012

Contact: Cordonnier-Pratt MM

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in

the Human Genome Center, University of Tokyo Institute of Medical

Science; plant material and RNA prepared at Texas A &amp; M University;

sequencing done in the Laboratory for Genomics and Bioinformatics;

University of Georgia. Sequence ends have been trimmed to exclude

vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sug5 (CTTCTGCTCTAAAGAGTCGG).

## FEATURES

source

Location/Qualifiers  
1. .1112  
/organism="Sorghum bicolor"  
/mol type="mRNA"  
/cultivar="IS3620C"  
/db xref="taxon:4558"  
/clone="ABAL\_14\_B04\_A012"  
/lab host="DH10B-T1 phage-resistant E. coli"  
/clone lib="Abscisic acid-treated seedlings"  
/note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from seedlings grown in hydroponic culture. After 12 days, medium was supplemented with 1 mM abscisic acid (ABA), while leaves were misted with a solution of 1 mM ABA. Roots and leaves were harvested after 3, 6, 12, and 24 hr and material from all time points was combined prior to RNA isolation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 112;

Best Local Similarity 100.0%; Pred. No. 1.8e+03; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 CGGCCCGCGCGCGACGA 434

Db 89 CGGCCCGCGCGCGACGA 71

## RESULT 30

BF822655

LOCUS

DEFINITION

CM2-RT0009-091200-605-b05 RT0009 Homo sapiens

ACCESSION

BF822655.1

VERSION

GI:12162819

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

Eukaryote; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 211)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags.

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

MEDLINE

10737800

PUBMED

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&amp;tl2=CM2-RT0009-

091200-605-b05&amp;tl3=2000-12-09&amp;tl4=1)

Seq primer: puc 18 forward

High quality sequence start: 24

High quality sequence stop: 158.

FEATURES  
source

Location/Qualifiers  
1. .211  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="RT0009"  
/note="Organ: kidney\_tumor; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 3.8%; Score 19; DB 10; Length 211;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 CATCTGCTGAAGCTGGTC 45  
|||||  
DB 134 CATCTGCTGAAGCTGGTC 152  
|||||

## RESULT 31

AI352333 219 bp mRNA linear EST 30-DEC-1998  
LOCUS qt19g03.x1 NCI CGAP GC4 Homo sapiens cDNA clone IMAGE:1948084 3'  
DEFINITION similar to SW:UROC\_RAT P55090 UROCORTIN PRECURSOR ; mRNA sequence.  
ACCESSION AI352333  
VERSION AI352333.1 GI:4089539  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 219)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbtp/image/image.html  
Seq primer: -40UP from Gibco.

FEATURES  
source

Location/Qualifiers  
1. .219  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1948084"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP GC4"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 219;

Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGAACCGCATC 30  
|||||  
DB 72 CGCCGAGCAGAACCGCATC 90  
|||||

## RESULT 32

AI631150 219 bp mRNA linear EST 14-DEC-1999  
LOCUS ts93c06.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2238826 3'  
DEFINITION similar to TR:077577 077577 UROCORTIN PRECURSOR ; mRNA sequence.  
ACCESSION AI631150  
VERSION AI631150.1 GI:4682480  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 219)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbtp/image/image.html  
Insert Length: 298 Std Error: 0.00  
Seq primer: -40UP from Gibco.

FEATURES  
source

Location/Qualifiers  
1. .219  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2238826"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP GC6"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGAACCGCATC 30  
|||||  
DB 72 CGCCGAGCAGAACCGCATC 90  
|||||

## RESULT 33

BF773138 220 bp mRNA linear EST 12-JAN-2001  
LOCUS CMA-IT0042-151200-605-b05 IT0042 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BF773138  
ACCESSION BF773138  
VERSION BF773138.1 GI:12121038





Seq primer: M13 Forward  
POLYA=No.

## FEATURES

source

Location/Qualifiers  
1. .346  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-Y0-acf-b-01-0-UI"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-Y0"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-Y0 library is a subtracted library derived from an individually-tagged normalized whole-eye (minus the lens) library. The driver for the subtraction consisted of a pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)  
TAG TISSUE=Eye  
TAG\_LIB=UI-R-Y0  
TAG\_SEQ=CAATG"

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 346;  
Best Local Similarity 100.0%; Pred. No. 2e+03; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0;  
QY 101 GCCAGCGTTCAGGAGGT 119  
|||||  
Db 149 GCCAGCGTTCAGGAGGT 131

## RESULT 36

BE415413

LOCUS

DEFINITION

BE415413 349 bp mRNA linear EST 24-JUL-2000  
clone MWL030.B02000309 ITEC MWL Wheat Root Library Triticum aestivum CDNA  
clone MWL030.B02, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Triticum aestivum (bread wheat)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.

REFERENCE

AUTHORS

Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T.,  
Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A.,  
Gustafson, P., Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J.,  
Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P.,  
Ogihara, Y., Pecchioni, N., Qaluset, C., Schuch, W., Selvaraj, G.,  
Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.  
International Triticaceae EST Cooperative (ITEC): Production of  
Expressed Sequence Tags for Species of the Triticeae  
Unpublished (2000)  
Contact: Warburton M  
Applied Biotechnology Center, CIMMYT

TITLE

JOURNAL

COMMENT

Appo. Postal 6-641, 06600 Mexico DF MEXICO  
Tel: 52-5-7269091 ext 1381  
Fax: 52-5-7267558/59  
Email: mwarburton@cgmnet.com  
International Triticaceae EST Cooperative (ITEC)  
http://wheat.pw.usda.gov/genome.

## FEATURES

source

Location/Qualifiers  
1. .349  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Atlas"  
/db\_xref="taxon:4565"  
/clone="MWL030.B02"  
/tissue\_type="root"  
/dev\_stage="8 day old"  
/clone\_lib="ITEC MWL Wheat Root Library"  
/note="Vector: pYES2 (Invitrogen); 0.5-1.5 Kbp average  
insert size."

## ORIGIN

Query Match 3.8%; Score 19; DB 10; Length 349;  
Best Local Similarity 100.0%; Pred. No. 2e+03; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0;  
QY 31 CTGCTGAAGCTGGTCCGCG 49  
|||||  
Db 83 CTGCTGAAGCTGGTCCGCG 101

## RESULT 37

BF941825

LOCUS

DEFINITION

BF941825 356 bp mRNA linear EST 30-MAR-2001  
nad3h05.x1 NCI CGAP Ov18 Homo sapiens cDNA clone IMAGE:3433569 3'  
similar to SW:UROC\_MOUSE P81615 UROCORIN PRECURSOR. [1] ;, mRNA  
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL

COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Christopher A. Moekaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento  
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 334.

Location/Qualifiers

FEATURES

source

1. .356  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3433569"  
/tissue\_type="fibrotheoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NCI CGAP\_Ov18"  
/note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTCGGAGCGCGCGACATTTTTTTTTTTT 3'];  
double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

## ORIGIN

Query Match 3.8%; Score 19; DB 10; Length 356;  
Best Local Similarity 100.0%; Pred. No. 2e+03;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 CGCCGAGCAGACCGCATC 30  
|||||  
Db 72 CGCCGAGCAGACCGCATC 90

## RESULT 38

CA646451

## LOCUS

CA646451 358 bp mRNA linear EST 23-NOV-2002  
wreln.pk0095.h11 wreln Triticum aestivum cDNA clone  
wreln.pk0095.h11 5' end, mRNA sequence.

## ACCESSION

CA646451

## VERSION

CA646451.1 GI:25224747

## KEYWORDS

EST.

## SOURCE

Triticum aestivum (bread wheat)

## ORGANISM

Triticum aestivum

## REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Triticum.

## AUTHORS

1 (bases 1 to 358)

## TITLE

Miao, G., Caraher, N. and Hanafey, M.K.

## JOURNAL

Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z., Miao, G., Caraher, N. and Hanafey, M.K.

## COMMENT

DuPont Wheat cDNA Sequence

Unpublished (2002)

## CONTACT

Contact: Scott V. Tingey

Crop Genetics

E. I. DuPont de Nemours and Company

1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA

Tel: 302-631-2602

Fax: 302-631-2607

Email: Scott.V.Tingey@USA.dupont.com

Seq primer: M13.

## FEATURES

source

1..358

Location/Qualifiers

/organism="Triticum aestivum"

/mol\_type="mRNA"

/db\_xref="taxon:4565"

/clone="wreln.pk0095.h11"

/tissue\_type="root"

/clone\_lib="wreln"

/note="Vector: pBluescript SK+; Site\_1: EcoRI; Site 2: XhoI; wheat (Triticum aestivum L.) root; normalized from wreln library"

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 358;

Best Local Similarity 100.0%; Pred. No. 2e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 CTGCTGAAGCTGTCGCG 49

|||||

Db 135 CTGCTGAAGCTGTCGCG 153

## RESULT 39

H24268

## LOCUS

H24268 379 bp mRNA linear EST 06-JUL-1995

IMAGE:52383 5' similar to gb:X63692 DNA (HUMAN); mRNA sequence.

## ACCESSION

H24268

## VERSION

H24268.1 GI:892963

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 379)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilton RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 1983

High quality sequence stops: 370 Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1983 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 370.

Location/Qualifiers

1..379

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="GDB:425319"

/db\_xref="taxon:9606"

/clone="IMAGE:52383"

/sex="female"

/dev\_stage="73 days post natal"

/lab\_host="DH10B (ampicillin resistant)"

/clone\_lib="Soares infant brain INIB"

/note="Organ: whole brain; Vector: Lafmid BA; Site\_1: Not I; Site 2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5,

AACGGAAGATTCCGCGCAGGAGATTTTTTTTTTTT 3'];

double-stranded cDNA was ligated to Hind III adaptors

(Pharmacia), digested with Not I and directionally cloned

into the Not I and Hind III sites of the Lafmid BA vector.

Library went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo."

Query Match 3.8%; Score 19; DB 14; Length 379;

Best Local Similarity 100.0%; Pred. No. 2.1e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 CATCTGCTGAAGCTGGTC 45

|||||

Db 297 CATCTGCTGAAGCTGGTC 315

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 379;

Best Local Similarity 100.0%; Pred. No. 2.1e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 CATCTGCTGAAGCTGGTC 45

|||||

Db 297 CATCTGCTGAAGCTGGTC 315

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 379;

Best Local Similarity 100.0%; Pred. No. 2.1e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 CATCTGCTGAAGCTGGTC 45

|||||

Db 297 CATCTGCTGAAGCTGGTC 315

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 379;

Best Local Similarity 100.0%; Pred. No. 2.1e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 CATCTGCTGAAGCTGGTC 45

|||||

Db 297 CATCTGCTGAAGCTGGTC 315

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 379;

Best Local Similarity 100.0%; Pred. No. 2.1e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 CATCTGCTGAAGCTGGTC 45

|||||

Db 297 CATCTGCTGAAGCTGGTC 315

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 379;

Best Local Similarity 100.0%; Pred. No. 2.1e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 CATCTGCTGAAGCTGGTC 45

|||||

Db 297 CATCTGCTGAAGCTGGTC 315

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 379;

Best Local Similarity 100.0%; Pred. No. 2.1e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 CATCTGCTGAAGCTGGTC 45

|||||

Db 297 CATCTGCTGAAGCTGGTC 315

```

PUBMED
COMMENT
8899548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LLNL (info@image.llnl.gov). IMAGE ID= 1791476
Seq primer: M13 Forward.
FEATURES
source
Location/Qualifiers
1..384
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-act-b-01-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-Y0"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; The UI-R-Y0
library is a subtracted library derived from an
individually-tagged normalized whole-eye (minus the lens)
library. The driver for the subtraction consisted of a
pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dt track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-Y0) was constructed as follows: PCR
amplified cDNA inserts from previous library clones from
which 3' ESTs had been derived were used as a driver in a
hybridization with the normalized whole-eye library in
the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-Y0
library. This procedure has been previously described
(Bonaldo, Lemmon and Soares, Genome Research 6: 791-806,
1996)"
ORIGIN
Query Match 3.8%; Score 19; DB 10; Length 384;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 101 GCCAGGCGTTCAGGAGGT 119
|||||
Db 212 GCCAGGCGTTCAGGAGGT 230
|||||

RESULT 41
CA614966 399 bp mRNA linear EST 23-NOV-2002
LOCUS
DEFINITION
wrl.pk149.h5 wrl Triticum aestivum cDNA clone wrl.pk149.h5 5' end,
mRNA sequence.
ACCESSION
CA614966
VERSION
CA614966.1 GI:25193277
KEYWORDS
EST.
SOURCE
Triticum aestivum (bread wheat)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Triticum.
1 (bases 1 to 399)
REFERENCE
Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
DuPont Wheat cDNA Sequence
TITLE
DuPont Wheat cDNA Sequence

JOURNAL
COMMENT
Unpublished (2002)
Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
FEATURES
source
Location/Qualifiers
1..399
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wrl.pk149.h5"
/tissue_type="root"
/clone_lib="wrl"
/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat (Triticum aestivum L.) root; 7 day old
seedling, light grown"
ORIGIN
Query Match 3.8%; Score 19; DB 14; Length 399;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 31 CTGCTGAAGCTGTCGCG 49
|||||
Db 14 CTGCTGAAGCTGTCGCG 32
|||||

RESULT 42
BE415414 410 bp mRNA linear EST 24-JUL-2000
LOCUS
DEFINITION
MWL030.B03000309 ITEC MWL Wheat Root Library Triticum aestivum cDNA
Clone MWL030.B03, mRNA sequence.
ACCESSION
BE415414
VERSION
BE415414.1 GI:9413260
KEYWORDS
EST.
SOURCE
Triticum aestivum (bread wheat)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Triticum.
1 (bases 1 to 410)
REFERENCE
Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T.,
Cloutier,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A.,
Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J.,
Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P.,
Ogihara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G.,
Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.
International Triticeae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)
Contact: Warburton M
Applied Biotechnology Center, CIMMYT
Apdo. Postal 6-641, 06600 Mexico DF MEXICO
Tel: 52-5-7269091 ext 1381
Fax: 52-5-7267558/59
Email: mwarburton@cnet.com
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.
FEATURES
source
Location/Qualifiers
1..410
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Atlas"
/db_xref="taxon:4565"
/clone="MWL030.B03"
/tissue_type="root"
/dev_stage="8 day old"
/clone_lib="ITEC MWL Wheat Root Library"
/note="Vector: pYES2 (Invitrogen); 0.5-1.5 Kbp average

```